

STIC-Biotech/ChemLib

73150

From: Chan, Christina  
Sent: Tuesday, August 13, 2002 10:44 AM  
To: Davis, Minh-Tam; STIC-Biotech/ChemLib  
Subject: RE: Rush search request for 09/697206

Please rush. Thanks Chris

-----Original Message-----

From: Davis, Minh-Tam  
Sent: Tuesday, August 13, 2002 10:37 AM  
To: Chan, Christina  
Subject: Rush search request for 09/697206

Please search in commercial database and in issued patent files:  
SEQ ID NO:1.

Thank you.

MINH TAM DAVIS

ART UNIT 1642, ROOM 8A01, MB 8E12  
305-2008

porter 10/94

Point of Contact:  
Mona Smith  
Technical Information Specialist  
CM1 6A01  
Tel: 308-3278

(1)

RECEIVED  
AUG 13 2002  
STIC/Biotech/ChemLib  
(STIC)

Searcher: M. Smith  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 8/13/02  
Date Completed: 8/14/02  
Searcher Prep/Review: 5  
Clerical: \_\_\_\_\_  
Online time: 7

TYPE OF SEARCH:  
NA Sequences: 1  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**

Point of Contact  
More Smith  
Technical Information Specialist  
OMI 8401  
Tel. 202-371-7070

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2002, 05:39:44 ; Search time 18384.3 seconds  
(without alignments)  
4080.749 Million cell updates/sec

Title: US-09-697-206A-1  
Perfect score: 3585  
Sequence: 1 cttttggatcactctgtg.....gctgtttaatctcaaaaaa 3585

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
-----					

1	2360	65.8	2731	6	AX098197
2	1962.8	54.8	2140	9	AK000178
3	1534.4	42.8	15933	2	AC009692
c	1449.8	40.4	91840	2	AC084253
5	1324.8	37.0	1624	9	AK001805
6	640.8	17.9	3406	6	AK025571
7	498.4	13.9	503	6	AX062424
8	498.4	13.9	503	6	AX062425
9	485.2	13.5	499	6	AX340160
c	434	12.1	434	6	AX351465
11	351	9.8	353	6	AX193256
c	350.4	9.8	377	6	AX335697
13	343.8	9.6	396	6	AX093263
14	343.8	9.6	396	6	AX193162
15	302.4	8.5	146089	2	AC108860
16	303.4	8.5	15933	2	AC009692
c	271	7.6	271	6	AX302732
18	237.2	6.6	5247	3	AF321979
c	228.4	6.4	171157	2	AC020779
c	228.4	6.4	196860	2	AC091121
c	223	6.2	294	6	AX098156
22	212.6	5.9	141289	9	AL449363
23	212.6	5.9	163569	2	AC023293
24	166	4.6	241	6	AX302731
25	156.2	4.4	230231	2	AC094385
c	154.6	4.3	164299	9	AC020978
27	132.2	3.7	195623	9	AP004218
c	127.4	3.6	120391	2	AC017146
c	127.4	3.6	173165	3	AC099005
c	127.4	3.6	192366	3	AC008369
c	127.4	3.6	258873	3	AE003810
32	118.2	3.3	339	6	AX261840
33	110.4	3.1	1433	9	AK025901
34	99.6	2.8	1611	3	AY052064
35	76.2	2.1	79	6	AX341241
36	76.2	2.1	79	6	AX341505
37	71.2	2.0	1910	10	AB022209
38	69.6	1.9	2121	10	BC018185
c	69.6	1.9	216425	2	AC084825
c	68.6	1.9	42778	3	CEZK1067
41	68	1.9	202152	2	AC095577
42	67.6	1.9	2365	10	BC005461
c	67.6	1.9	88871	10	MMU58105
44	66.4	1.9	1902	9	HUMHFP
45	66.4	1.9	2168	9	BC015580

ALIGNMENTS

RESULT	1				
AX098197	AX098197	Sequence 109 from Patent WO0118542.	2731 bp	DNA	linear PAT 30-MAR-2001
LOCUS	AX098197	Sequence 109 from Patent WO0118542.			
DEFINITION	AX098197	Sequence 109 from Patent WO0118542.			
ACCESSION	AX098197	Sequence 109 from Patent WO0118542.			
VERSION	AX098197.1	GI:13515284			
KEYWORDS	human.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Lee, J., Thompson, P., and Lillie, J.				
TITLE	Identification, assessment, prevention, and therapy of ovarian cancer				
JOURNAL	Patent: WO 0118542-A 109 15-MAR-2001; Millennium Predictive Medicine, Inc. (US)				
FEATURES	Location/Qualifiers				
source	1..2731				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
BASE COUNT	825 a	562 c	549 g	795 t	
ORIGIN					



Qy 2744 gctacatgtagttagtaattcttagttgaactgaatgaatattggtggtctcatatgt 2803  
 Db 1855 GCTACATGAGTAGTAACTTCTAGTTGAACGTAAATGAATATGCTGGCTTCATATGT 1914  
 Qy 2804 attattttatattgtaatttttttattgattgaggttgacttttaataagagaattcc 2863  
 Db 1915 ATTATTTTATATTGATGCTTTTTCATTTATTGATGGTTTGGACCTTAATAAGAGAAATCC 1974  
 Qy 2864 atagtttttaataatccagagtgagacaatttgaacagtgatttcttagaatacaataca 2923  
 Db 1975 ATAGTTTTTAATATCCAGAGTGAACAAATTTGAACAGTGTATCTAGAGAAACAATACA 2034  
 Qy 2924 ctaactgaacagagtgtaatttatatatattatgatagaccttaaacctttttccctcta 2983  
 Db 2035 CTAAGTGAACAGAGTGAATGCTTATATATATATATATATAGCTTTAAACCTTTTCCCTCTA 2094  
 Qy 2984 atgctttaactgctcaataattataaacttttaaacgcataggaactatagtcagcatcta 3043  
 Db 2095 ATGCTTTAACTGTCACAAATAATTATAACCTTTTAAAGCATAGGACTATAGTCAGCATGCTA 2154  
 Qy 3044 gactgagaggttaaacactgatgcaattagacaggtactgatgctgcagtggtttaacac 3103  
 Db 2155 GACTGAGAGGTAAACACTGATGCAATTTAGAACAGTACTGATGCTGTCAGTGTTAACAC 2214  
 Qy 3104 tatgtttagctgtgtttatgctataaaagtgcataattagacactagctagctagctgc 3163  
 Db 2215 TATGTTTATGCTGTGTTTATGCTATATAAAGTGCAATATTAGACACTAGCTAGTACTGCTGC 2274  
 Qy 3164 ctcattgaactccaaagaaacaggatttcatatgaattgcattgaattggaattttctct 3223  
 Db 2275 CTCATGTAACCTCCAAAGAAACAGGATTTTCAATTAAGTGCATTTGAATGTGGCTATTCTCT 2334  
 Qy 3224 aagttaactcatattgtcttcttgctgaatgaatgcctgcagctgctgctgctat 3283  
 Db 2335 AAGTACTCATATTTGCTTGTGTAATGCAATGCCGTGCAGATTTATGTGGCTGCTAT 2394  
 Qy 3284 tttattttctgtgcttaacttaacaccttaaggagaggaagcaaacattttccctcttca 3343  
 Db 2395 TTTTATTTTCTGTGCTACTTTAAACACTTTAAAGGAGAGCAACATTTCCCTTCTTCA 2454  
 Qy 3344 gctgactggcaatggccctttaactgcaataggaagaaaaaaaggggttggtgtaaa 3403  
 Db 2455 GCTGACTGGCAATGGCCCTTTAACTGCAATAGGAAGAAAAAAGGGTTGTTGTGAAA 2514  
 Qy 3404 attggtgataactggcacttaagatcgaagaaatttctgtatacttgatgcttcaaga 3463  
 Db 2515 ATTGGTGATACACTGGCACTTAAGATCGAAGAAATTTCTGTATACCTTGATGCTTTAAGA 2574  
 Qy 3464 tgccaaagctgcccagaagcttgaaagactttaagatagggcagtaattgcttactacaat 3523  
 Db 2575 TGCCAAAGCTGCCAAAGCTCTGAAGACTTTAAAGATAGSCAGTAATGCTTACTACAAT 2634  
 Qy 3524 actactgagttttgttagagtttaacatttgataataaaacttgccctgttaactcctaaa 3583  
 Db 2635 ACTACTGAGTTTTGTAGAGTTTAACTTTGATATAATAAACTTGCCCTTTTAACTCCAAA 2694  
 Qy 3584 aa 3585  
 Db 2695 AA 2696  
 RESULT 2  
 LOCUS AK000178 2140 bp mRNA linear PRI 22-FEB-2000  
 DEFINITION Homo sapiens cDNA FLJ20171 fis, clone COL09761.  
 ACCESSION AK000178  
 VERSION AK000178.1 GI:7020093  
 KEYWORDS oligo capping; fis (full insert sequence).  
 SOURCE Homo sapiens colon cDNA to mRNA, clone\_lib:COL clone:COL09761.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (sites)

AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.  
 NEDO human cDNA sequencing project  
 unpublished (2000)  
 REFERENCE 2 (bases 1 to 2140)  
 AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.  
 Direct Submission  
 Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@leims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)  
 NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing; Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).  
 FEATURES  
 Location/Qualifiers  
 1..2140  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="COL09761"  
 /tissue\_type="colon"  
 /clone\_lib="COL"  
 /note="cloning vector pME18SFL3"  
 58..1134  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="BAA90992.1"  
 /db\_xref="GI:7020094"  
 /translation="MPEYLNFEKSSSVRYCASQVEDMGNIILLIMISEYPNHRSDPE RVNKFESGTCRSMELIDDDVVRARGLPWQSSDDIARFKGLNIAKGAALCLNAQ GRNGEALRVFSEEHRLDALQRKHHMGTRYIEVKATGEDFLKTAGTSNEVAQFL SKENQIVRMNRLGPFATAEVVAFFGQHGILFTVYDGRPTGDAFLVLA FVPTVNRDLCIRLGLPYAATIEDILDFLGEFATDINTHGVHVLNHNQGRPSGDAFLQ MKSADRAFMAAQKCHKKHKGQIC"  
 BASE COUNT 606 a 481 c 464 g 589 t  
 ORIGIN  
 Query Match 54.8%; Score 1962.8; DB 9; Length 2140;  
 Best Local Similarity 94.2%; Pred No: 0;  
 Matches 2120; Conservative 0; Mismatches 7; Indels 123; Gaps 3;

Qy 394 gaattcaagaatgttgccctggttcacctgatattgacaaactggacgttgccacaatg 453  
 Db 1 GAATTCAGAAATGTTGCCCTGGTTCCACTGATATTGACAAACTGGACGTTGCCACAATG 60  
 Qy 454 acagagatttttaatttttgagaagagtagttcagtcctctcgatatggagcctcacaagt 513  
 Db 61 ACAGAGATTATTAATTTTGAAGAAGTAGTTTCAGTCTCTCGATATGGAGCCTCTCAAGTT 120  
 Qy 514 gaagatatggggaataataatttttagcaatgatttcagagccttataatcacaggtttcca 573  
 Db 121 GAAGATATGGGAATATATAATTTTAGCAATGATTTTCAGAGCCTTATATACAGGTTTCA 180  
 Qy 574 gatccagagagtggaattacaagtgttgaaagtggaaacttcagacaagatggaacttatt 633  
 Db 181 GATCCAGAGAGTGAATTAACAAGTTTGAAAGTGGAACTTCAGACAAGATGGAACCTTATT 240  
 Qy 634 gatataaacacgttagtcaggggcagaggtttaccatggcagcttcagatcagaatatt 693  
 Db 241 GATGATAACACCGTAGTCAGGGGCACGAGGTTTACCATGGCAGTCTTCAGATCAAGATATT 300  
 Qy 694 qcaagattttcaaggaactcaatttgccaaaggaggtgcagcactttgtcctaagtct 753  
 Db 301 GCAAGATCTTTCGAAGGACTCAATATTTGCCAAGGAGGTGCAGCACTTTGTCTGAATGCT 360  
 Qy 754 cagggtcgaaggaaacggagaagcgtgtagttgttaggtttgtaagtgaggagcaccgagaccta 813



Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A.,  
 Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A.,  
 Cooke, P., DeArcellano, K., Depayre, E., Devon, K., Dewar, K.,  
 Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C.,  
 Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,  
 Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L.,  
 Karatas, A., Lehotsky, J., Lieu, C., Locke, K., Macdonald, P.,  
 Marquis, N., McEwan, P., McGuck, A., McKernan, K., McLaughlin, J.,  
 Meldrim, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,  
 Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,  
 Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,  
 Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,  
 Testaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,  
 Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.  
 Direct Submission  
 Submitted (28-AUG-1999) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Sep 10, 2000 this sequence version replaced gi:7547130.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L2303  
 Center clone name: lil\_D13  
 ----- Summary Statistics  
 Sequencing vector: M13; M77815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 132474 bases at least Q40  
 Consensus quality: 144968 bases at least Q30  
 Consensus quality: 150672 bases at least Q20  
 Insert size: 182000; agarose-fp  
 Insert size: 156233; sum-of-contigs  
 Quality coverage: 3.0 in Q20 bases; agarose-fp  
 Quality coverage: 3.5 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 38 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1 532: contig of 532 bp in length  
 \* 533 632: gap of 100 bp  
 \* 633 1684: contig of 1052 bp in length  
 \* 1685 1784: gap of 100 bp  
 \* 1785 2846: contig of 1062 bp in length  
 \* 2847 2946: gap of 100 bp  
 \* 2947 4075: contig of 1129 bp in length  
 \* 4076 4175: gap of 100 bp  
 \* 4176 5462: contig of 1287 bp in length  
 \* 5463 5562: gap of 100 bp  
 \* 5563 6755: contig of 1193 bp in length  
 \* 6756 6855: gap of 100 bp  
 \* 6856 8009: contig of 1154 bp in length  
 \* 8010 8109: gap of 100 bp  
 \* 8110 9599: contig of 1490 bp in length  
 \* 9600 9699: gap of 100 bp  
 \* 9700 10876: contig of 1177 bp in length  
 \* 10877 10976: gap of 100 bp  
 \* 10977 12499: contig of 1523 bp in length  
 \* 12500 12599: gap of 100 bp  
 \* 12600 13603: contig of 1004 bp in length  
 \* 13604 13703: gap of 100 bp  
 \* 13704 15007: contig of 1304 bp in length  
 \* 15008 15107: gap of 100 bp

15108 16547: contig of 1440 bp in length  
 \* 16548 16647: gap of 100 bp  
 \* 16648 17884: contig of 1237 bp in length  
 \* 17885 17984: gap of 100 bp  
 \* 17985 19796: contig of 1812 bp in length  
 \* 19797 19896: gap of 100 bp  
 \* 19897 21294: contig of 1398 bp in length  
 \* 21295 21394: gap of 100 bp  
 \* 21395 22935: contig of 1541 bp in length  
 \* 22936 23035: gap of 100 bp  
 \* 23036 26495: contig of 3460 bp in length  
 \* 26496 26595: gap of 100 bp  
 \* 26596 27849: contig of 1254 bp in length  
 \* 27850 27949: gap of 100 bp  
 \* 27950 45165: contig of 17216 bp in length  
 \* 45166 45265: gap of 100 bp  
 \* 45266 47264: contig of 1999 bp in length  
 \* 47265 47364: gap of 100 bp  
 \* 47365 51079: contig of 3715 bp in length  
 \* 51080 51179: gap of 100 bp  
 \* 51180 54251: contig of 3072 bp in length  
 \* 54252 54351: gap of 100 bp  
 \* 54352 58073: contig of 3722 bp in length  
 \* 58074 58173: gap of 100 bp  
 \* 58174 61307: contig of 3134 bp in length  
 \* 61308 61407: gap of 100 bp  
 \* 61408 65572: contig of 4165 bp in length  
 \* 65573 65672: gap of 100 bp  
 \* 65673 69628: contig of 3956 bp in length  
 \* 69629 69728: gap of 100 bp  
 \* 69729 73243: contig of 3515 bp in length  
 \* 73244 73343: gap of 100 bp  
 \* 73344 77727: contig of 4384 bp in length  
 \* 77728 77827: gap of 100 bp  
 \* 77828 83613: contig of 5786 bp in length  
 \* 83614 83713: gap of 100 bp  
 \* 83714 89358: contig of 5645 bp in length  
 \* 89359 89458: gap of 100 bp  
 \* 89459 95302: contig of 5924 bp in length  
 \* 95303 95482: gap of 100 bp  
 \* 95483 102960: contig of 7478 bp in length  
 \* 102961 103060: gap of 100 bp  
 \* 103061 112213: contig of 9153 bp in length  
 \* 112214 112313: gap of 100 bp  
 \* 112314 125685: contig of 13372 bp in length  
 \* 125686 125785: gap of 100 bp  
 \* 125786 142030: contig of 16245 bp in length  
 \* 142031 142130: gap of 100 bp  
 \* 142131 157955: contig of 15825 bp in length  
 \* 157956 158055: gap of 100 bp  
 \* 158056 159933: contig of 1878 bp in length.  
 \*  
 \* Location/Qualifiers  
 \* 1. .159933  
 \* /organism="Homo sapiens"  
 \* /db\_xref="taxon:9606"  
 \* /chromosome="15"  
 \* /map="15"  
 \* /clone="RP11-111D13"  
 \* /clone\_lib="RPC1-11 Human Male BAC"  
 \* 1. .532  
 \* /note="assembly\_fragment  
 \* clone\_end:SP6  
 \* vector\_side:left"  
 \* 633. .1684  
 \* /note="assembly\_fragment"  
 \* 1785. .2846  
 \* /note="assembly\_fragment"  
 \* 2947. .4075  
 \* /note="assembly\_fragment"  
 \* 4176. .5462  
 \* /note="assembly\_fragment"  
 \* 5563. .6755  
 \* /note="assembly\_fragment"

FEATURES

source

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature





```

RESULT      4
AC084253
LOCUS       AC084253 Homo sapiens clone RP11-498C11, WORKING DRAFT SEQUENCE, 5 unordered
DEFINITION pieces.
ACCESSION   AC084253
VERSION     AC084253.3 GI:13123246
KEYWORDS    HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 91840)
AUTHOR      Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens, clone RP11-498C11
JOURNAL
REFERENCE
AUTHORS
2 (bases 1 to 91840)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bada,F., Boguslavskiy,L.,
Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
Lamazares,R., Landers,T., Lechoczky,J., Levine,R., Lieu,C., Liu,G.,
Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Melorim,J., Meneus,L., Mihoval,T., Mlenga,V.,
Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
O'Donnell,P., O'Neill,D., Olivivar,T.M., Oliver,J., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Travers,M., Trigglio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (18-OCT-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 25, 2001 this sequence version replaced gi:11597050.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11484
Center clone name: 498_C_11
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 90064 bases at least Q40
Consensus quality: 90922 bases at least Q30
Consensus quality: 91180 bases at least Q20
Insert size: 85000; agarose-fp
Insert size: 91440; sum-of-contigs
Quality coverage: 11.1 in Q20 bases; agarose-fp
Quality coverage: 10.3 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```

```

* be preserved.
* 1 29791: contig of 29791 bp in length
* 29792 29891: gap of 100 bp
* 29892 31022: contig of 1131 bp in length
* 31023 31122: gap of 100 bp
* 31123 57388: contig of 26266 bp in length
* 57389 57488: gap of 100 bp
* 57489 91339: contig of 33851 bp in length
* 91340 91439: gap of 100 bp
* 91440 91840: contig of 401 bp in length.
FEATURES
            Location/Qualifiers
            source
            1..91840
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="RP11-498C11"
            /clone_lib="RPCI-11 Human Male BAC"
            misc_feature
            1..29791
            /note="assembly_fragment"
            clone_end:SP6
            vector_side:left"
            misc_feature
            29892..31022
            /note="assembly_fragment"
            misc_feature
            31123..57388
            /note="assembly_fragment"
            misc_feature
            57489..91339
            /note="assembly_fragment"
            misc_feature
            91440..91840
            /note="assembly_fragment"
            clone_end:rf
            vector_side:right"
BASE COUNT 26234 a 17891 c 18577 g 28684 t 454 others
ORIGIN
Query Match 40.4%; Score 1449.8; DB 2; Length 91840;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 1534; Conservative 0; Mismatches 10; Indels 10; Gaps 7;
QY 2038 agtgttgaagatgtatggtgatcttgaaacctccagacacacagaaacctcttagcaaa 2097
|||||
Db 22622 AGTGTGGAAGATGATGGTGATCTTGAAACCTCCAGACACAGAAACCTCTTAGCAAA 22681
QY 2098 ttccaggggaagtgtgtctacactcagctcagctcagctatttttcagcaaaccttgattggacaaa 2157
|||||
Db 22682 TTCAGGGGAAGTTGTTCTACACTCAGGCTGCAGTATTTTCAGCAACCTGATTGGACAAA 22741
QY 2158 cgggcctgtgcttatcttttggaggagtgaaataatttgagctagtgaaagccaaatcgt 2217
|||||
Db 22742 CGGGCCTGTGCTTATCTTTTGGTGGAGTGAAAAAATTTGAGCCAGTGAAGCCAAATCGT 22801
QY 2218 aactacagcaagcagcatgcagcatcctgtgctcttctgtgattgcaaataggcattta 2277
|||||
Db 22802 AACTTACAGCAGCAGCATGCAGCATACCTGGCTCTCTTCTGCTGATTGCAATAGGCATTTA 22861
QY 2278 aaatgtgaatttggaatcagatgtctccacttccagtttaaagtggcatcatagtggt 2337
|||||
Db 22862 AAATGTGAATTTGGAATCAGATGCTCCATTAATTTAAAGTGSCATCATAGGCGT 22921
QY 2338 ttccatagttttaaagtcttggaataaaaact---ccaccagtgctctaccatctccc-accatg 2394
|||||
Db 22922 TTCCTAAGTTTTTAAGTCTTGGATAAAAAATNCCNACCAGTGTCTACCATCTCCNACCATG 22981
QY 2395 aactcgttaagggaagcttcattttgtatatctccgcctctttctctctcttctccctgt 2454
|||||
Db 22982 AACNCTGTTAAGGAAGCTTCAATTTTCGTATATATCCCGCTCTTTTCTCTTCATTTCCCTGT 23041
QY 2455 ctctgcataatcatgcctctcttctgtaagttaattcaagcataaagatctcttgaataataaa 2514
|||||
Db 23042 CTTCTGCATAATCATGCTTCTTCTGCTAAGTAATTTCAAGCATAAAGATCTTGTGAATAATAA 23101
QY 2515 atcacatcttaggagagaagaataaaattgttatttcccagtgctcttctggccatgatgat 2574
|||||
Db 23102 ATCACAATCTTAGGAGAAAGAAATAAAATTTGTTATTTTCCCAAGTCTCTTGGCCATGATGAT 23161

```

QY	2575	atcttatgataaaacaaataataattttaaacacacctgaagataaaattagaagaattg	2634	DEFINITION	Homo sapiens cDNA FLJ10943 fis, clone OVARC1001360.
Db	23162	ATCTTATGATTAAACAAATAATTAATTTTAAACACCTGAAGATATATTAGAAGAAATTG	23221	ACCESSION	AK001805
QY	2635	tgaacctcc-acaaaacatacaaaagttaaagaatttggaatctcttttctcagcaggtatc	2693	VERSION	AK001805.1 GI:7023305
Db	23222	TGCACCTCCNACAAAACATACAAAGCTTTAAAGATTTGGATCTTTTCTCAGCAGTATC	23281	KEYWORDS	oligo capping; fis (full insert sequence).
QY	2694	agttgtaataatgaattaggcgccaaaatgcaaaacgaaataagcagctacatgta	2753	SOURCE	Homo sapiens ovarian cancer cDNA to mRNA, clone_lib:OVARC1 clone:OVARC1001360.
Db	23282	ACTGTAAATAATGAATTAGGCGCCAAATGCAAAACGAAAAATGAAGCAGCTACATGTA	23341	ORGANISM	Homo sapiens
QY	2754	gttagtaatttctagttgaaactgtaattgaatatattggtgcttcacatagtattattat	2813	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Db	23342	GTTAGTAATTTCTAGTTTGAACGTAAATCAATATTGTGGCTTCATATGATATTATTAT	23401	AUTHORS	1 (sites) Isozai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H., Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K., Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masubo,Y., Ninomiya,K. and Iwayanagi,T.
QY	2814	atgtgaacttttccattatgattggattggacttttaataagagaataatccatgatttta	2873	TITLE	NEDO human cDNA sequencing project
Db	23402	ATTGTACTTTTTTCATTTATGATGGTTTGGACTTTAATAAGAGAAATTCCTAGTTTTTA	23461	JOURNAL	Unpublished (2000)
QY	2874	atccccagaagtgcagaacttgaacagtgtaattctagaaaaatacacactaactgaac	2933	REFERENCE	2 (bases 1 to 1624)
Db	23462	ATATCCCGAAGCTGAGACAAATTTGACACGTGTATTTCTAGAAAAACAATACACTAAGTGAAC	23521	AUTHORS	Isogai,T. and Otsuki,T.
QY	2934	agaagtgaatgccttatatatattatgatagcccttaaccccttttcccttaatgccttaac	2993	TITLE	Direct Submission
Db	23522	AGAAGTGAATGCTATATATATTATGATAGCCTTAAACCTTTTTCCTCTAATGCCTTAAC	23581	JOURNAL	Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases, Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)
QY	2994	tgcaaaataattaaacctttttaaagcataggaactatagtcacatgctagactgagagg	3053	COMMENT	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, - University of Tokyo.
Db	23582	TGTCAAATAATTAACCTTTTAAAGCATAGGACTATAGTCAGCATGCTAGACTGAGAGG	23641	FEATURES	Location/Qualifiers
QY	3054	taaacctgatgcaattagAACAGGtactgatgctgcagtggtttaaactatgtttagc	3113	Source	1..1624 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="OVARC1001360" /clone_lib="OVARC1" /tissue.type="ovarian cancer" /note="Cloning vector: pME18SFL3"
Db	23642	TAAACACTGATGCAATTAGAACAGGTACTGATGCTGTCAGTGTTTAACTATGTTTAGC	23701	BASE COUNT	498 a 320 c 288 g 518 t
QY	3114	tggtgttatgctataaaagtgcataattagacacactagctagtagtgc--tgctcatgta	3171	ORIGIN	
Db	23702	TGTGTTTATGCTATAAAGTGAATATTAGACACTAGTAGTACTGCGTGCCTCATGTA	23761	Query Match	37.0%; Score 1324.8; DB 9; Length 1624;
QY	3172	actcc--aaagaaacaggatttcattaaagtgcattgaattggtatattctctaagtta	3229	Best Local Similarity	99.5%; Pred. No. 6.9e-308;
Db	23762	ACTNCCAAAGNAAACAGGATTTCAATTAGTGATTTGAATGTGGCTATTTCTCTAAGTTA	23821	Matches 1329; Conservative	0; Mismatches 7; Indels 0; Gaps 0;
QY	3230	ctcatattgt-ctttgtcttgaatgcaatgccgtgcagatttatgagctgctattttta	3288	QY	2038 agtgtttgaaagatgtagtgatcttgaacccctccagacacaaagaaacctctagcaaa
Db	23822	CTCATATTGTTNCCTTGCTTGAATGCAATCCCGTGAGATTTTATGTGGCTGCTATTTTA	23881	Db	289 AGTGTTTGAAAGATGTATGGTGATCTTGAAACCTCCAGACACAGAAACCTTAGCAAA
QY	3289	ttttc-tgtgcattactttaacaccttaaggaggagaaacatttccttcttcagctg	3347	QY	2098 ttcaggggaagttgtctacacacagctgcagatttttcagcaaaacttgatggacaaa
Db	23882	TTTTTCNTGTGCATTACTTTTAACACCTTAAAGGGAGAGCAACATTTCCCTTCCTCAGCTG	23941	Db	349 TTCAAGGGGAAGTTTGTCACACTCAGCTCAGTATTTTCAGCAAACTTGATTTGGACAAA
QY	3348	actggaatggccctttaactcaatagagagaaaaaagggttggtaaaattg	3407	QY	2158 cgggcctgtgccttatcttttgggtgagtgtaaaaaatttgagctagtgaaagccaaatcgt
Db	23942	ACTGGCAATGGCCCTTTAACTGCAATAGGAAGAAAAAAGGTTTGTGGAATTTG	24001	Db	409 CGGCGCTGTGCCCTATCTTTTGGTGAGTGAAAAATTTTGAGCCAGTGAAGCCAAATCGT
QY	3408	gtgataactggcacttaagatcgaagaagaatttctgtatctgtatgccttaagatgcc	3467	QY	2218 aacttacagaacgacatgcagcatcacctgcgtctcttctgtgattgcaaataggcattta
Db	24002	GTGATACTGGCACTTAAGATCGAAAGAAATTTCTGTATACTTGTATGCGCTTAAGATGCC	24061	Db	469 AACTTACAGCAAGCAGCATGCAGCATACCTTGGCTCTTTGCTGATTGCAAAATAGGCATTTA
QY	3468	caaaagctgccccaaagctctgaagacatttaagataggcagtaaatgtctactacaatacta	3527	QY	2278 aaatggaatttggaatcagatgtccattcttacttccagttaaagtggcatactagtgct
Db	24062	CAAAAGCTGCCAAAGCTCTGAAAGACTTTTAAGATAGGCAGTAATGCTTACTACAATACTA	24121	Db	529 AAATGTGAATTTGGAATFCAGATGTCCTCAATACTTCCAGTTAAGTGGCATCATAGCCGT
QY	3528	ctgaatttttgaggttaacatttgataataaaacttgcctgttttaactcaaa	3581	QY	2338 ttccctaagttttaagctcttgataaaaaactcccaccagtgctctaccatctccacatgaac
Db	24122	CTGAGTTTTTGTAGAGTTACATTTTGATAATAAACTTGCCCTGTTTAATCTCAA	24175	Db	589 TTCTTAAGTTTTTAAGTCTTGATATAAAACTCCACCAGTGTCTACCATCTCCACCATGAAC
LOCUS	AK001805	1624 bp	mrna	linear	PRI 22-FEB-2000
RESULT	5				





FEATURES CORIXA CORPORATION (US)  
source Location/Qualifiers  
1. .503  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 188 a 69 c 78 g 168 t  
ORIGIN

Query Match 13.9%; Score 498.4; DB 6; Length 503;  
Best Local Similarity 99.8%; Pred. No. 4.5e-109;  
Matches 499; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2572 gatattcttattgattaaaaaacaaattaaatttttaaacacacctgaagaataattagaagaaa 2631  
Db 1 GATATCTTATGATTAAAAACAATAATTAATTTTAAAAACACCTGAAGATATATTAGAAGAAA 60

QY 2632 ttgtgcacccctccacaaacatacaaaagttttaaaagtgttgatcttttctcagcaggtga 2691  
Db 61 TTGTGCACCCCTCCACAAAACATACAAAAGTTTAAAGTTTGGATCTTTTCTCAGCAGGTA 120

QY 2692 tcaagtgttaataatgaattaggggcccaaaatgcgaaacgaaaaatgaagcagctacatg 2751  
Db 121 TCAGTTGTAAATGAATTAGGGGCCAAAATGCAAAACGAAAATGAAGCAGCTACATG 180

QY 2752 tagttagtaattcttagtttgaactgttaattgaattattgtggtcttcataatgattatttt 2811  
Db 181 TAGTTAGTAAATTTCTAGTTTGAACGTGAATTCGAATATTGTGCTTCATATGTATTATTTT 240

QY 2812 atattgtacttttttcattattgattggttgaccttaataagagaaaattccatagtttt 2871  
Db 241 ATATTGTACTTTTTTCATTATTATTGATGGTTTGGACTTTTAATAAGAGAAAATTCATAGTTTT 300

QY 2872 taatcccccagaagtgcagacaatttgaacagtgattctcagaaaaacatacaactga 2931  
Db 301 TAATATCCAGAGTGCAGACAATTTGAACAGTGTAATCTAGAAAACAATACACTAAGTGA 360

QY 2932 acagaagtgaatgcttatatatattatgatagccttaaacctttttccctcaatgcctta 2991  
Db 361 ACAGAAGTGAATGCTTATATATATTATGATAGCCTTAAACCTTTTCCCTTAATGCCTTA 420

QY 2992 actgtcaataataattaaaccttttaaaagcataggactatagtcagcatgctagactgaga 3051  
Db 421 ACTGTCAAAATAATTAAACCTTTTAAAGCATAGGACTAGTACAGCATGCTAGACTGAGA 480

QY 3052 ggttaaacactgatgcaatta 3071  
Db 481 GGTAAACACTGATGCAATTA 500

RESULT 8  
LOCUS AX062425 503 bp DNA linear PAT 24-JAN-2001  
DEFINITION Sequence 52 from Patent WO0100828.  
ACCESSION AX062425  
VERSION AX062425.1 GI:12540300  
KEYWORDS human.  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 503)  
AUTHORS Wang,T., Bangur,C.S., Lodes,M.J., Fanger,G.R., Vedvick,T.S.,  
Carter,D., Retter,M.W. and Mannion,J.  
TITLE Compositions and methods for the therapy and diagnosis of lung  
cancer  
JOURNAL Patent: WO 0100828-A 52 04-JAN-2001;  
CORIXA CORPORATION (US)  
FEATURES  
source Location/Qualifiers  
1. .503  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 188 a 69 c 78 g 168 t

Query Match 13.5%; Score 485.2; DB 6; Length 499;  
Best Local Similarity 99.6%; Pred. No. 6.8e-106;  
Matches 496; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

ORIGIN

Query Match 13.9%; Score 498.4; DB 6; Length 503;  
Best Local Similarity 99.8%; Pred. No. 4.5e-109;  
Matches 499; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2572 gatattcttattgattaaaaaacaaattaaatttttaaacacacctgaagaataattagaagaaa 2631  
Db 1 GATATCTTATGATTAAAAACAATAATTAATTTTAAAAACACCTGAAGATATATTAGAAGAAA 60

QY 2632 ttgtgcacccctccacaaacatacaaaagttttaaaagtgttgatcttttctcagcaggtga 2691  
Db 61 TTGTGCACCCCTCCACAAAACATACAAAAGTTTAAAGTTTGGATCTTTTCTCAGCAGGTA 120

QY 2692 tcaagtgttaataatgaattaggggcccaaaatgcgaaacgaaaaatgaagcagctacatg 2751  
Db 121 TCAGTTGTAAATGAATTAGGGGCCAAAATGCAAAACGAAAATGAAGCAGCTACATG 180

QY 2752 tagttagtaattcttagtttgaactgttaattgaattattgtggtcttcataatgattatttt 2811  
Db 181 TAGTTAGTAAATTTCTAGTTTGAACGTGAATTCGAATATTGTGCTTCATATGTATTATTTT 240

QY 2812 atattgtacttttttcattattgattggttgaccttaataagagaaaattccatagtttt 2871  
Db 241 ATATTGTACTTTTTTCATTATTATTGATGGTTTGGACTTTTAATAAGAGAAAATTCATAGTTTT 300

QY 2872 taatcccccagaagtgcagacaatttgaacagtgattctcagaaaaacatacaactga 2931  
Db 301 TAATATCCAGAGTGCAGACAATTTGAACAGTGTAATCTAGAAAACAATACACTAAGTGA 360

QY 2932 acagaagtgaatgcttatatatattatgatagccttaaacctttttccctcaatgcctta 2991  
Db 361 ACAGAAGTGAATGCTTATATATATTATGATAGCCTTAAACCTTTTCCCTTAATGCCTTA 420

QY 2992 actgtcaataataattaaaccttttaaaagcataggactatagtcagcatgctagactgaga 3051  
Db 421 ACTGTCAAAATAATTAAACCTTTTAAAGCATAGGACTAGTACAGCATGCTAGACTGAGA 480

QY 3052 ggttaaacactgatgcaatta 3071  
Db 481 GGTAAACACTGATGCAATTA 500

RESULT 9  
LOCUS AX340160 499 bp DNA linear PAT 10-JAN-2002  
DEFINITION Sequence 407 from Patent WO0196388.  
ACCESSION AX340160  
VERSION AX340160.1 GI:18136141  
KEYWORDS human.  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Jiang,Y., Harlocker,S.L. and Secrist,H.  
TITLE Compositions and methods for the therapy and diagnosis of colon  
cancer  
JOURNAL Patent: WO 0196388-A 407 20-DEC-2001;  
CORIXA CORPORATION (US)  
FEATURES  
source Location/Qualifiers  
1. .499  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 155 a 90 c 118 g 134 t 2 others  
ORIGIN

Query Match 13.5%; Score 485.2; DB 6; Length 499;  
Best Local Similarity 99.6%; Pred. No. 6.8e-106;  
Matches 496; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

```
Qy 351 actattacctgaatgcttctattctcttttttttgatctctcgaagaattcaagaatgttg 410
Db 1 ACTATTACCTGAATGCTTCTATTCTCTTTTGTATCTTCGAAAGAAATCAAGAAATCTTG 60
Qy 411 ccttggttcaactgatattgacaaactggacgtttgcacaaatgacagagatttttaaat 470
Db 61 CCCGTTGTCACCTGATATTGACAAACTGGACGTTGCCACAAATGACAGAGTA-TTAAATTT 119
Qy 471 tgagaagtagtctcagtcctcgatatggagcctctcaagtgtgaagatatggggaatat 530
Db 120 TGAGAAGAGTAGTTCAGTCTCTCATATGGAGCCTCTCAAGTTCGAAGATATGGGGAATAT 179
Qy 531 aatttgacaaatgatttcagagccttataatcacaggtttttcagatccagagagagtgaa 590
Db 180 AATTTTGAATGATTTTCAGAGCCTTATAATCACAGGTTTTCAGATCCAGAGAGAGTGAA 239
Qy 591 ttacaagtttgaaagtggaaacttgcagaagatggaacttattgatataacacccgtagt 650
Db 240 TTACAAGTTTGAAGTGGAACTTCGACCAAGATGGAACCTTATGTATGATGATAACACCGTAGT 299
Qy 651 cagggcacagagtttaccatggcagcttcagatcaagatatgtcaagattcttcaaaag 710
Db 300 CAGGCACGAGGTTTACCATGGCAGTCTTCAGATCAAGATATTCAGAGATCTTCAAGG 359
Qy 711 actcaatttgcgaaggagggtgcagcactttgtctgaatgctcaggtgcgaagaaacgg 770
Db 360 ACTCAATATTGCCAAGGAGGTCGAGCAGTCTGTCTGAATGCTCAGGTCGGAAGAACGG 419
Qy 771 agaagctctggttaggtttgaatgaggagcaccgagacccagacttagcactacagagacaa 830
Db 420 AGAAGCTCTGGTTAGGTTTGAAGTGAGGACCGCAGACCTAGCAGTACAGAGGACAA 479
Qy 831 acatcacatggggaccgcg 848
Db 480 ACATCACATGGGGACCCD 497

RESULT 10
AX351465/c
LOCUS AX351465 434 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 212 from Patent WO0196390.
ACCESSION AX351465
VERSION AX351465.1 GI:18616812
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
TITLE Jiang,Y., Hepler,W.T., Clapper,J.D., Wang,A. and Secrist,H.
JOURNAL Compositions and methods for the therapy and diagnosis of colon
cancer
PATENT: WO 0196390-A 212 20-DEC-2001;
CORIXA CORPORATION (US)
FEATURES
source
1. .434
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 112 a 114 c 73 g 135 t
ORIGIN

Query Match 12.1%; Score 434; DB 6; Length 434;
Best Local Similarity 100.0%; Pred. No. 1.4e-93;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 470 ttgagaagtagtctcagtcctcgatatggagcctctcaagtgtgaagatatggggaata 529
Db 434 TTGAGAAGAGTAGTTCAGTCTCTCATATGGAGCCTCTCAAGTTCGAAGATATGGGGAATA 375
Qy 530 taatttgaagaatgatttcagagccttataatcacaggtttttcagatccagagagagatga 589
Db 374 TAATTTTAGCAATGATTTTCAGAGCCTTATAATCACAGGTTTTCAGATCCAGAGAGAGTGA 315
```

```
Qy 590 attacaagtttgaagtggaaacttgacagcaagatggaaactttatgatgataaacaccgtag 649
Db 314 ATTACAGTTTGAAGTGGAACTTCGACAGAGATGGAACTTATTGATGATTAACACCGTAG 255
Qy 650 tcagggcacaggtttaccatggcagcttctcagatcaagatatgtcaagattcttcaaaag 709
Db 254 TCAGGGCACAGGTTTACCATGGCAGTCTTCAGATCAAGATATTGCAAGATTTCTTCAAAG 195
Qy 710 gactcaatattgcaaggaggtgcagcactttgtctgaatgctcaggttcgaaggaacg 769
Db 194 GACTCAATATTGCCAAGGAGGTCAGCAGTCTGTCTCAATGCTCAGGTCGAAGGAACG 135
Qy 770 gagaagctctggttaggtttgaagtgcagcaccgagacccagactagcactacagagggcaca 829
Db 134 GAGAAGCTCTGGTTAGGTTTGAAGTGAGGAGCACCAGAGCCTAGCAGTACAGAGGCACA 75
Qy 830 aacatcacatgggaccgcggtatattgaggtttacaagaacacgggtgaagattctccta 889
Db 74 AACATCACATGGGACCGGTATATTGAGGTTTACAAAGCAACAGGTGAAGATTTCCTTA 15
Qy 890 aaattgctggtggt 903
Db 14 AATTGCTGGTGGT 1

RESULT 11
AX193256
LOCUS AX193256 353 bp DNA linear PAT 15-AUG-2001
DEFINITION Sequence 823 from Patent WO0149716.
ACCESSION AX193256
VERSION AX193256.1 GI:15211207
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Xu,J., Lodes,M.J., Secrist,H., Benson,D.R., Meagher,M.J.,
Stolk,J.A., King,G.E., Wang,T. and Jiang,Y.
TITLE Compounds for immunotherapy and diagnosis of colon cancer and
methods for their use
JOURNAL Patent: WO 0149716-A 823 12-JUL-2001;
CORIXA CORPORATION (US)
FEATURES
source
1. .353
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 122 a 46 c 54 g 129 t 2 others
ORIGIN

Query Match 9.8%; Score 351; DB 6; Length 353;
Best Local Similarity 99.4%; Pred. No. 1.3e-73;
Matches 351; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2663 aaaagtttgatcttttctcagcaggtatcagttgtaataataatgaggggcaaaa 2722
Db 1 AAAAGTTTGGATCTTTTCTCAGCAGGTCATCAGTTGTAATAATGAATTAGGGGCCAAAA 60
Qy 2723 tgcaaacgaaaaatgaagcagctcacatgtagttagtaattctagttgaaactgtaatt 2782
Db 61 TGCAAAACGAAAAATGAAGCAGCTACATGTAGTAGTAATTTCTAGTTGAACGTGAATT 120
Qy 2783 gaattatggcttcacatgtattatttatgtacttttttcattatgattggttg 2842
Db 121 GAATATTGGCTTCAATATGATTATTATTATTTGTTACTTTTTCATTATTATGATGNTTG 180
Qy 2843 gactttaaagaagaaatcccatagtttttaatatccagagtgagacaaatttgaacag 2902
Db 181 GACTTTAATAGACAAATTCCTATAGTTTAAATATCCAGAGTGAGACAAATTTGAACAG 240
Qy 2903 tgtattctagaaaaacaatacacactaacctgaacagagtgatgcttatatatattatgata 2962
```

```
Db 241 TGTATTCTAGAAAACATACATACTAACTGAACAGAGTGAATGCTTATATATATATNATA 300
|||||
QY 2563 gccttaaaccttttctctaaagccttaactgtcctaaataattataacotttt 3015
|||||
Db 301 GCCTTAAACCTTTTCTCTAATGCCCTTAAGTGTCAATAATATATAACCTTTT 353
|||||

RESULT 12
AX335697/c AX335697 377 bp DNA linear PAT 09-JAN-2002
LOCUS
DEFINITION Sequence 6206 from Patent WO0194629.
ACCESSION AX335697
VERSION AX335697.1 GI:18126416
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
Horigan,S., Soppet,D.R. and Weaver,Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 6206 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
Location/Qualifiers
source 1..377
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 122 a 70 c 66 g 118 t 1 others
ORIGIN

Query Match 9.8%; Score 350.4; DB 6; Length 377;
Best Local Similarity 98.9%; Pred. No. 1.8e-73;
Matches 373; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 3205 tgaatgtgatat-tcttaagtctactcatattgt-ctttgtctgaatgcaatgccgt 3262
|||||
Db 377 TGAATGTGGATATTCCCTTAAGTTACTCATATATGNCCTTTGCTTGAATGCAATGCCGT 318
|||||
QY 3263 gcagattatgagcgtcgtatttttttctgtgcattactttaacaccttaaaaggag 3322
|||||
Db 317 GCAGATTATGAGCTGCTATTTTATTCTGTGCATTACTTTAACACCTTAAAGGGAG 258
|||||
QY 3323 aagcaaacattctcttcagctgactgcaatggccctttaactgcaataggaagaaa 3382
|||||
Db 257 AAGCAACATTTCTCTTCAGCTGACTGGCAATGGCCCTTTAACTGCAATAGGAAGAAA 198
|||||
QY 3383 aaaaaaagggttctgtgaaaaattggtataactggcacttaagatcgaagaatttc 3442
|||||
Db 197 AAAAAAAGGTTCTGTGAAAATTGGTGATACTTAAGTGGCACTTAAGATCGAAAATAATTC 138
|||||
QY 3443 tgtatacttgatgccttaagatgccaaagctgccaaagctctgaaagactttaagata 3502
|||||
Db 137 TGTATATTGATGCTTAAGATGCCAAAGCTGCCAAAGCTCTGAAAGACTTTAAAGATA 78
|||||
QY 3503 ggcagtaatgcttactacaataactactgagttttttagagtttaacatttgataaaaa 3562
|||||
Db 77 GGCAAGTAATGCTTACTACAATACTACTGAGTTTTTGTGAGAGTTAAACATTTGATAATAAA 18
|||||

QY 3563 ctgctcgttttaactc 3579
|||||
Db 17 CTGCGCTGTTTATCTC 1
|||||

RESULT 13
AX093263
LOCUS
DEFINITION Sequence 81 from Patent WO0118046.
ACCESSION AX093263
VERSION AX093263.1 GI:13509712
```

```
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Xu,J. and Stolk,J.A.
TITLE Ovarian tumor sequences and methods of use therefor
JOURNAL Patent: WO 0118046-A 81 15-MAR-2001;
CORIXA CORPORATION (US)
FEATURES
Location/Qualifiers
source 1..396
/organism="Homo sapiens"
/db_xref="taxon:9606"
misc_feature 1..396
/notes="n = A,T,C or G"
BASE COUNT 101 a 96 c 88 g 101 t 10 others
ORIGIN

Query Match 9.6%; Score 343.8; DB 6; Length 396;
Best Local Similarity 93.7%; Pred. No. 7.1e-72;
Matches 374; Conservative 0; Mismatches 22; Indels 3; Gaps 2;

QY 1186 gcagctgaagttcagcaggtgctgaatcgattctcctcgccctctcattccactcca 1245
|||||
Db 1 GCAGCTGAAGTTTCAGCAGGTGCTGAATCGATTCTCTCGGCCCTCTCATTTCCACTTCCA 60
|||||
QY 1246 accctccattattccagttactacctcagcaattttgtgccccctacaaatgtagagac 1305
|||||
Db 61 ACCCTCCCATTTATCCAGTACTACCTCAGCAATTTGTGCCCTTACAAATGTTAGAGAC 120
|||||
QY 1306 tgtatacgcttcagagctcttcctctatgcagccacaattgagacatccctggatttcctg 1365
|||||
Db 121 TGTATAGCCCTTCAGGCTTCCTCTATGCAGCCACAATTGAGGACATCTCGCATTTCTG 180
|||||
QY 1366 gggaggtccacagatatctactcatggggttcacatgggtttgaatcacaccagggc 1425
|||||
Db 181 GGGAGGTTGCCACAGATATTCGTACTCATGGGTTTCACATGGTTTGAATCACACGGGN 240
|||||
QY 1426 cgccatcaggagatgcctttatccagatgaagctctcgacagagcattatggctgca 1485
|||||
Db 241 CCGCCATCAGAGATGCTTTATCCAGATGAAGTCTCGGACAGANCAATTTATGGCTGCA 300
|||||
QY 1486 cagaagtgtcataaaaaacatgaagcacagatatgttgaagctcttcagtgctcagct 1545
|||||
Db 301 CAGAAAGTGGCATAAAAAACATGAAGCACAGATATGTTGAAG--TTTTTCAGTGTGAGCT 358
|||||
QY 1546 gagagatgaacttgtgttaatgggggggcactttaaat 1584
|||||
Db 359 GANGANA-GAACATTGNNGTANNNGGGGNACTTTAAAT 396
|||||

RESULT 14
AX193162
LOCUS
DEFINITION Sequence 729 from Patent WO0149716.
ACCESSION AX193162
VERSION AX193162.1 GI:15211113
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Xu,J., Lodes,M.J., Secrist,H., Benson,D.R., Meagher,M.J.,
Stolk,J.A., King,G.E., Wang,T. and Jiang,Y.
TITLE Compounds for immunotherapy and diagnosis of colon cancer and
methods for their use
JOURNAL Patent: WO 0149716-A 729 12-JUL-2001;
CORIXA CORPORATION (US)
FEATURES
Location/Qualifiers
source 1..329
```

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 97 a 58 c 63 g 108 t 3 others
ORIGIN

Query Match 9.0%; Score 322.8; DB 6; Length 329;
Best Local Similarity 98.5%; Pred. NO. 7.9e-67;
Matches 324; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3016 aaagcatagactatagtcagcatgctagcagaggttaaacactgagcgaattagaac 3075
Db 1 AAAGCATAGACTATAGTCAGCATGCTAGCAGAGGTAAACACTGATGCAATTAGAAC 60

QY 3076 aggtactgctgctcagtggttaaacactgcttagctggtttatgctataaaagtgc 3135
Db 61 AGGTACTGCTGCTCAGTGTAAACACTATGTTAGCTGTTTATGCTATATAAGTGC 120

QY 3136 aaattagacactagctagctgctcgtctgtaacctccaaagaaaacaggatttcac 3195
Db 121 AATATTAGACACTAGCTAGCTGCTGCTCATGTAACTCCAAGAAAAACAGGATTTCA 180

QY 3196 taagtgcattgagtggtattctctcaagttactcatattgctccttgccttgatgca 3255
Db 181 TAAGTGCATTGAGTGGCTATTCTCTAAGTTACTCATATTGCTCTTGGCTGAATGCA 240

QY 3256 atgcccgcagatttatgagcgtctattttatttctgtcctactttaaaccctta 3315
Db 241 ATGCCGCGAGATTATGTCGCTGCTATTTTATTTTCGTGNCATTTACTTTAACACCTTA 300

QY 3316 aaggggaagcaaacatttcctcttcag 3344
Db 301 AAGGAGAAGCAAAACATTTCTCTTCTTCAG 329

RESULT 15
AC108860
LOCUS
DEFINITION Homo sapiens chromosome 8 clone RP11-22C11 map 8, WORKING DRAFT
SEQUENCE, 9 ordered pieces.
ACCESSION AC108860
VERSION AC108860.1 GI:18450124
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 8, clone RP11-22C11
JOURNAL Unpublished
REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhaltier,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepeil,Y., Colangelo,M., Collins,S., Collimore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
```

Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (31-JAN-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L21833  
----- Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 143556 bases at least Q40  
Consensus quality: 144302 bases at least Q30  
Consensus quality: 144739 bases at least Q20  
Insert size: 145289; sum-of-contigs  
Quality coverage: 12.5 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 9 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 21231: contig of 21231 bp in length  
\* 21232 21331: gap of 100 bp  
\* 21332 22852: contig of 1521 bp in length  
\* 22853 22952: gap of 100 bp  
\* 22953 26035: contig of 3083 bp in length  
\* 26036 26135: gap of 100 bp  
\* 26136 39308: contig of 13173 bp in length  
\* 39309 39408: gap of 100 bp  
\* 39409 52135: contig of 12727 bp in length  
\* 52136 52235: gap of 100 bp  
\* 52236 69562: contig of 17327 bp in length  
\* 69563 69662: gap of 100 bp  
\* 69663 85425: contig of 15763 bp in length  
\* 85426 85525: gap of 100 bp  
\* 85526 127682: contig of 42157 bp in length  
\* 127683 127783: gap of 100 bp  
\* 127783 146089: contig of 18307 bp in length.  
-----  
FEATURES  
Location/Qualifiers  
1..146089  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="8"  
/map="8"  
/clone="RP11-22C11"  
/clone.lib="RPC1-11 Human Male BAC"  
1..21231  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left  
21332..22852  
/note="assembly\_fragment"  
22953..26035  
/note="assembly\_fragment"  
26136..39308  
/note="assembly\_fragment"  
39409..52135  
/note="assembly\_fragment"  
52236..69562  
/note="assembly\_fragment"



misc\_feature 69663..85425  
/note="assembly\_fragment"  
misc\_feature 85526..127682  
/note="assembly\_fragment"  
misc\_feature 127783..146089  
/note="assembly\_fragment  
clone\_end:T7

BASE COUNT 40241 a 30457 c 30839 g 43752 t 800 others  
ORIGIN

Query Match 8.5%; Score 303.4; DB 2; Length 146089;  
Best Local Similarity 99.7%; Pred. No. 4.8e-62;  
Matches 304; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 901 ggtacttccaataggtagccagtttctctccaagaaaataagtcattgttcgcatg 960  
Db 118645 GGTACTTCCAATAGGTAGGCCAGTTTCTCTCCAAGGAAAATCAAGTCATTGTCCGCATG 118704  
QY 961 cgggggctccctttcacggccacagctgaagaagtgtggccctctttggacagcattgc 1020  
Db 118705 CGGGGGCTCCCTTTACGGCCACAGCTGAAGAAGTGGTGGCCTTCTTTGGACAGCATTGC 118764  
QY 1021 cctattactggggaaagggaaggaatcctctttgtcacctaccagatggttaggccaaca 1080  
Db 118765 CCTATTACTGGGGAAAGGAAGGCATCCTCTTTGTCACTACCCAGATGGTAGGCCAACA 118824  
QY 1081 ggggacgttttgcctctttgctgtgaggaatatcacagaatgcggttgagggaagcat 1140  
Db 118825 GGGGACGCTTTTGTCTCTTTGCCCTGTGAGGAATATCACAGAAATGCGTTGAGGAAGCAT 118884  
QY 1141 aaagacttgttgggtaaaagatacatgttaactcttcaggagcacagcagctgaagttcag 1200  
Db 118885 AAAGACTTGTGGTAAAGATACATTGAACCTTTCAGGAGCACAGCAGCTGAAGTTTCAG 118944  
QY 1201 caggt 1205  
Db 118945 CAGGT 118949

Search completed: August 15, 2002, 14:11:07  
Job time: 30683 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2002, 04:42:29 ; Search time 8772.8 seconds  
(without alignments)  
5515.523 Million cell updates/sec

Title: US-09-697-206A-1

Perfect score: 3585

Sequence: 1 ctttttggatcactgctgg.....gctgtttaatctcaaaaa 3585

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpi:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1062.8	29.6	1137	9	AL545672
2	986.8	27.5	1038	9	AL545648
3	769.6	21.5	777	10	BG182742
4	742.2	21.3	791	10	BG191000
5	741.6	20.7	786	10	BG216627
6	715.2	19.9	798	10	BG193725
7	666	18.6	768	10	BG209741
8	653.2	18.2	736	10	BG286508
9	637	17.8	782	10	BG217686
10	633.8	17.7	936	10	BG291485
11	629.6	17.6	913	10	BF557167
12	627.6	17.5	960	10	BG287429
13	619.8	17.3	819	10	BG184305
14	618	17.2	904	10	BE870387
15	615.8	17.2	890	10	BG260588
16	609.6	17.0	801	10	BF555187
17	608.4	17.0	819	10	BG218557

18	604.4	16.9	644	10	BG195224
19	603.4	16.8	646	10	BG217687
20	594	16.6	839	10	BG203925
21	593	16.5	804	10	BG188511
22	583	16.3	908	10	BI452459
c 23	581.2	16.2	628	10	BE710214
c 24	580.8	16.2	596	10	BF001941
c 25	560.6	15.6	729	10	BI694471
c 26	558	15.6	645	9	AW861027
c 27	557.6	15.6	667	9	AV698922
c 28	557.4	15.5	559	10	BM127059
c 29	553.6	15.4	594	9	AW363149
c 30	551.2	15.4	646	9	AW563406
c 31	550.2	15.3	574	9	AW382909
c 32	547.4	15.3	574	10	BG289507
c 33	545.2	15.2	650	10	BG403930
c 34	545	15.2	545	9	AI990751
c 35	544.8	15.2	548	9	AI923923
c 36	543	15.1	563	9	AI961474
c 37	537.4	15.0	542	9	AI813884
c 38	530.8	14.8	534	9	AA843844
c 39	530.6	14.8	865	10	BI852872
c 40	525.8	14.7	530	9	BE219683
c 41	525.4	14.7	639	9	AV694407
c 42	524.4	14.6	528	10	BI963927
c 43	522.8	14.6	526	9	AI301132
c 44	522.4	14.6	525	9	AI690278
c 45	520.8	14.5	524	9	AI963119

#### ALIGNMENTS

RESULT 1  
AL545672 1137 bp mRNA linear EST 16-FEB-2001  
LOCUS AL545672 LTI\_NFL006\_PL2 Homo sapiens cDNA clone CS0DI009YD20 5  
DEFINITION prime, mRNA sequence.  
ACCESSION AL545672  
VERSION AL545672.1 GI:12878154  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1137)  
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
JOURNAL Contact: Genoscope  
COMMENT Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
FEATURES  
Location/Qualifiers  
1..1137  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0DI009YD20"  
/clone\_lib="LTI\_NFL006\_PL2"  
/tissue\_type="placenta"  
/note="vector: pcwvSPORT 6; Site\_1: NotI; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched, double-stranded cDNA was digested with Not I and  
cloned into the Not I and Eco RV sites of the pcwvSPORT 6  
vector. Library was normalized. Library was constructed by  
Life Technologies. Contact : Feng Liang Life Technologies,  
a division of Invitrogen 9800 Medical Center Drive  
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371  
Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com"

BASE COUNT 279 a 297 c 266 g 287 t 8 others  
ORIGIN

```

Query Match      29.6%; Score 1062.8; DB 9; Length 1137;
Best Local Similarity 98.0%; Pred. No. 9.7e-206;
Matches 1103; Conservative 8; Mismatches 1; Indels 14; Gaps 3;

Qy 753 tcagggctgaaggaacgagaagctctgttagtttgaagtgaagagaccacgt 812
Db 1 TCAGGGTCGAAGGAACGAGAGCTCTGGTTAGTTTGTAACTKAGGAGCACCAGACCT 60

Qy 813 agcaactacagaggcacaacaatcacatgggacccgtatatattgagtttacaagaac 872
Db 61 AGCACTACAGAGGCACAAACATCACATGAGGACCCGCTATATTGAGGTTTACAAGCAAC 120

Qy 873 aggtgaagattctcttaaaattgctgtgttaactccaatgaagtagccccagttctctc 932
Db 121 AGGTGAAGATTCTCTTAATGCTGGTGTACTTCCAATGAGGTAGCCAGTTCTCTC 180

Qy 933 caaggaatacaagctattgttcgcatgaggggctcccttccacgagctgaaga 992
Db 181 CAAGGAAATCAAGTCATG-TGCGATGCGGGGCTCCCTTTCACGGCCACAGCTGAAGA 239

Qy 993 agtgtggcctctttgacagcaattccctattactgggggaaaggaagcattccttt 1052
Db 240 AGTGGTGGCCTCTTTCGACAGCATTCCTTACTTACTGGGGAAAGMAGSCATCCCTCTT 299

Qy 1053 tgcacctaccagatggttaggccaacagggagcgttttctctcttgcctgtgagga 1112
Db 300 TGTCACCTACCCAGATGGTAGGCCAACAGGGAGCGCTTTTKCTCTCTTTCCTGTAKGA 359

Qy 1113 atatgcagaatgctgtgaggaagcaataagacttgtgggtaaaagatacatggaact 1172
Db 360 ATATGCAGAAATCGTTGAGGAAGCATAAAGACTTGTGGGTAAABATACATGAAT 419

Qy 1173 ctacagagacagcagctgaagttcagaggtgctgaatcgattctcctggccctct 1232
Db 420 CTTCAGGAGCACAGCAGTGAAGTTCAGCAGGTGCTGAATCGATTCCTCGGGCCCTCT 479

Qy 1233 cattccacttccaccctccatttccagttactactaccctcagcaatttggccccctac 1292
Db 480 CATTCACCTTCCAACCCCTCCCATTTATCCAGTACTACCTCAGCMATTTGTGCCCCCTAC 539

Qy 1293 aaatgttagactgtatagccttcaggtcttccctatgcagcccaaataggagacat 1352
Db 540 AAATGTTTAGACTGTATACGCCCTTCGAGCTTTCCTATGCAGCCACAATGAGGACAT 599

Qy 1353 cctggatttctggggagctgcacagatattcgtactcattgaggttccatgggttt 1412
Db 600 CCTGGATTCTCTGGGGAGTTCGCCACAGATATTCGTACTCATGGGGTTCACATGGTTTT 659

Qy 1413 gaatcacaggccgcgcacatcagagatgccttttccagatgaagtctgcggacagc 1472
Db 660 GAATCACAGGGCGCCCATCAGGAGATGCTTTATCCAGATGAAGTCTCGGACAGAGC 719

Qy 1473 attatggctgcagagaagtgtctataaaaaaacatgaagagacagatagtgtgaagcttt 1532
Db 720 ATTTATGGCTGCACAGAAGTGTCTATAAAAAAACATGAAGACACAGATATGTGTAAGTCTT 779

Qy 1533 tcagtgttcagctgagagatgaacttgttaagtgggggcaacttaaatcgaataag 1592
Db 780 TCAGTGTTCAGCTGAGGAGTGAATCTTGTGTTAATGGGGGCACTTTAAATCGAAATGG 839

Qy 1593 cttaaccaccacg-----ccatgcctgtctcctccctccctacacatttccagc 1640
Db 840 CTTATCCCCACCGCATGTAAAGTTTACCATGCTGTCTCTCTCCCTCCCTACACATTTCCAGC 899

Qy 1641 tccgtgtcagttatttccacagaagctgcatttaccagccctctgtgattttgaatcc 1700
Db 900 TCCGTCTGCAGTTTATTCCTACAGAAGTGCCTATTTACAGCCCTCTGTGTGATTTTGAATCC 959

Qy 1701 acgagcactgcagccctccacagcgtactaccagcagggcactcagctcttcatgaacta 1760
Db 960 ACGAGCACTGCAGGCCCTCCACAGCGTACTACCCAGCAGGCACTCAGCTCTTCATGAATTA 1019

```

```

Qy 1761 cacagcgtactatccccagccccccaggttcgcttaagtagtcttgctacttccctacagc 1820
Db 1020 CACAGCGTACTATCCAGAGCCCCCAGGTCGCTTAATAGTCTTGGCTACTTCCCTACAGC 1079

Qy 1821 tgctaacttagcgtgtcctccacagcctggcacggtgggtcaga 1866
Db 1080 TGCTAAATCTTAGCGGTGTCCT-CACAGCCTGGCAGCGTGGTCTAGA 1124

RESULT 2
ALS45648/c
LOCUS      ALS45648 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI009YD20 3
DEFINITION prime, mRNA sequence.
ACCESSION AL545648
VERSION    AL545648.1 GI:12878130
KEYWORDS   EST.
SOURCE      human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1038)
AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES             source
     source           1..1038
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DI009YD20"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life technologies. Contact : Feng Liang Life technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"

BASE COUNT      322 a 173 c 164 g 351 t 28 others
ORIGIN
Query Match      27.5%; Score 986.8; DB 9; Length 1038;
Best Local Similarity 96.3%; Pred. No. 2.7e-190;
Matches 993; Conservative 24; Mismatches 13; Indels 1; Gaps 1;

Qy 2446 ttctcgtcttctgcataatcatgctcttctgtaagtaattcaagcataagctcttg 2505
Db 1030 TTCCCTGTCTCTGTCATATCATGCTCTCTTCTGTAAGTAATTCAAGCATAAGATCTKGG 971

Qy 2506 aataataaaatcacaaatcttaggagaaagaataaaattgtatttcccagctctctggc 2565
Db 970 AATAATAAAATCAAAATCTTAGGAGAAAGTTTAAATTTGTTATTTTCCACGCTCTCTGGC 911

Qy 2566 catgatgatctctatgataaaacaaataaaattttaaaccctgaagataaattag 2625
Db 910 CATGATCATATCTTATGATTAATAAACAAATTAATTTTAAACACCTGAAGATATATTAG 851

Qy 2626 aagaaattgtgcacccctccacaaacatacaaaagttaaagtgttgatctttcttcag 2685
Db 850 AAGAAATTTGTGCACCCCTCCCAACAAACATACAAAGTTTAAAGTTTGGATCTTTTCTCAG 791

Qy 2686 caggtatcaattgtataataatgaattagggcccaaaatgcaaaaacgaataaagaagcagc 2745
Db 790 CAGGTATCAGTTGTAAATAATGAATTAGGGGCCAAAATGCAAAAACGAAAAACGAAGCAGC 731

```

```
Qy 2746 tacatgtagttagtaattcttagttgaactgttaattgaattatgttggttcacatgtat 2805
|||||
Db 730 TACATGTAGTTAGTAATTCTTAGTTGAACGTGAATTGAATATGTGGCTTCATATGTAT 671

Qy 2806 tatattatattgtacttttttcattatattgtagtggttggaactttaataagagaattccat 2865
|||||
Db 670 TATTTTATATTGTACTTTTTTTTCATTATTGATGGTTTGGACTTTAATAAGAGAAATTCAT 611

Qy 2866 agtttttaataatccagaggaagtgagacaatttgaacagtgattcttagaaaaacaatacact 2925
|||||
Db 610 AGTKTTTAATATCCGAGAAGTGAGACAATTTGAACAGGTGTATCTPAGAAAACAATACACT 551

Qy 2926 aactgaacagagaagtgaatgcgttatatatattatgatagcctttaaacctttttccctcctaat 2985
|||||
Db 550 AACTGAACAGAGAAGTGAAGAGCTBATATATATATGATAGCCTTAAACMATTTTCTCTWAAT 491

Qy 2986 gcccttaactgcacaataattaaaccttttaagacataggaactatagtcagcatgctaga 3045
|||||
Db 490 GCCTTAACTGACAAATAATTATAACCTTKKTAAGCANTAGGACTATAGTCAGCATGCTAGA 431

Qy 3046 ctgagagtgaaacactgatgcaattagacacaggtactgatgctgcagtggttttaacacta 3105
|||||
Db 430 CTGAGAGTAAACACTGATGCAATTAGACAGGTACTGATGCTGTCAGTGKWTACACTA 371

Qy 3106 tgtttagctgtgtttatgcgtataaaagtgcgaattattagacactagctagctactgcct 3165
|||||
Db 370 TGTNTAGCNGTGTATTATGCTATATAAARNVCAATATTAGACACTAGCTAGTACSGCGCSB 311

Qy 3166 catgtaactccaaagaaaaacaggatttcatttaagtgcattgaattgagatttctcttaa 3225
|||||
Db 310 CATGTAACBCCAAAGAAAACAGGATKTCATTAAAGTGCAATKGAATGTGCBATBTCTTAA 251

Qy 3226 gttaactaatgtctcttgcgtgaatgcaatgcagtcaggtactgatgctgcagtggtttta 3285
|||||
Db 250 GTBACTCATATGTCCTTTGCBKNAATGCAATGCCGTGCAGATTATATGTGGCTGCTATT 191

Qy 3286 ttattttctgtcattactttaacaccttaaaaggagagagaacaaacttccttcttcagc 3345
|||||
Db 190 TTAATTTTC-GTGCATTACTTTAAACACCTTAAAGGGAGAAGCAAAACATTTCCTTCTTCAGC 132

Qy 3346 tgactggcaatggccctttaactgcgaataggaagaaaaaaaggtttgtgtgaaaaat 3405
|||||
Db 131 TGACTGGCAATGGCCCTTTAACTGCAATAGGAAGAAAAAAGGTTTGTGTGAAAAAT 72

Qy 3406 tgggtgaactggcacttaagatcgaaagaaattctgtactatgcttgaagtgccttaagtgc 3465
|||||
Db 71 TGGTGATAAATGGGCATTAAGATCGAAAGAAATTTCTGTATACTTGTATGCTTAAGATG 12

Qy 3466 cccaaagctgc 3476
|||||
Db 11 CCCAAAGCTGM 1

RESULT 3
BG182742
LOCUS BG182742 777 bp mRNA linear EST 21-APR-2001
DEFINITION RST1619 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG182742
VERSION BG182742.1 GI:13704429
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 777)
Harrington,J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith
,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher
,J., Danzig,J. and Ducar,M.
Creation of genome-wide protein expression libraries using random
activation of gene expression
```

```
JOURNAL Nat. Biotechnol. 19 (5), 440-445 (2001)
MEDLINE 21227151
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave., Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 552.
FEATURES
Location/Qualifiers
source 1..777
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_lines="HT1080"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."
BASE COUNT 251 a 139 c 136 g 251 t
ORIGIN
Query Match 21.5%; Score 769.6; DB 10; Length 777;
Best Local Similarity 99.5%; Pred. No. 3.8e-146;
Matches 772; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 2038 agtgttgaaagatgatatggtagcttgaaacctccagacacaaagaactctctagcaaa 2097
Db 2 AGTGTGTTGAAAGATGATGGTGATCTTGAACCTCCAGACACAAAGAAACTTCTAGCAAA 61

Qy 2098 ttcagggaagtgtctacactcagctcaggtattttcagaacactgattgagacaaa 2157
Db 62 TTCAGGGGAAGTTTGCTACACTCAGGCTGCAGTATTTTCAGCAACTTGATGGACAAA 121

Qy 2158 cgggcctgtgccttatcttttggtgagtgaaataatttgagctagtgaaagccaaatcgt 2217
Db 122 CGGGCTGTGCTTATCTTTTGGCGGAGTGAAAAAATTTGAACTAGTGAAGCCAAATCGT 181

Qy 2218 aacttaacggaagcagcatgacgataccctggctcttctgctgattgcaaaataggattta 2277
Db 182 AACTTACAGCAAGCAGCATGCGAGCATACCTGGCTCTTTGCTGATTGCCAAATAGGCATTTA 241

Qy 2278 aaatgtgaattgggaatcagatgtctccattacttccagttaaagtggcatacattggt 2337
Db 242 TAATGTGAATTTGGAAATCAGATGCTCTCCATTACTTCCAGTTACAGTGGCATTATAGTGT 301

Qy 2338 ttccctaagttttaagctcttggtgataaaaaactccaccagtgctaccatctccaccatgaac 2397
Db 302 TTCCCTAAGTTTAAAGCTTTGGATAAAAACTCCACCAGTGTCTTACCATTCTCCACCATGAAC 361

Qy 2398 tctgttaagggaagcttcattttgttatattcccgctctttctcttcttcttccctgtctt 2457
Db 362 TCTGTTAAGGAAGCTTCATTTTGTATATTTCCCGCTCTTTTCTCTCAATTTCCCTGTCTT 421

Qy 2458 ctgcataatcatgcctctcttgctaagtaattcaagcataagatcttggaataataaaatc 2517
Db 422 CTGCATAATCATGCCCTTCTTGGCTAAGTAATTCACACATAAGATCTTGGAAATTAATAATC 481

Qy 2518 acaatcttaggagagaagaataaaattgttattttcccgagctctcttggccatgatgatc 2577
Db 482 ACAATCTTAGGAGAAAGAAATAAAATTTGTTATTTTCCCAGTCTCTTGGCCATGATCATATC 541

Qy 2578 ttatgattaaaaacaataataattttaaacaacctgaagataaattagaagaattgtgc 2637
Db 542 TTATGATTAAAAACAATAATTAATTTTAAACACACCTTGAAGATAAAATTAGAAGAAATTTGTC 601

Qy 2638 accctccacaacaatacaaaagttaaagtttgagatcttttctcagcaggtatcagtt 2697
Db 602 ACCCTCCACAAAACATACAAAAGTTTAAAGTTTGGATCTTTTCTTCAGCAGGTATCAGTT 661
```

```
QY 2698 gtaataatgaattagggcccaaaatgcaaaacgaaagaaagcagctacatgtagtta 2757
|||||
Db 662 GTAATAATGAATTAGGGGCCAAATGCAAAACGAAAAATGAAGCAGCTACATGACTTA 721

QY 2758 gtaattctagttgaactgtaattgaattgttggttcctcatgtatttttat 2813
|||||
Db 722 GTAATTTCTAGTTGTAATGTAATGTAATGTTGGCTTCATATGTAATTTAT 777

RESULT 4
BG191000          791 bp      mRNA      linear      EST 21-APR-2001
LOCUS             RST10214 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
DEFINITION
ACCESSION          BG191000
VERSION            BG191000.1 GI:13712815
KEYWORDS           EST.
SOURCE             human.
ORGANISM           Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS            Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Costanzo,D., McElligott,K., Boozer,S., Mays,R., Smith
,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher
,J., Danzig,J. and Ducar,M.
TITLE              Creation of genome-wide protein expression libraries using random
activation of gene expression
JOURNAL            Nat. Biotechnol. 19 (5), 440-445 (2001)
MEDLINE
COMMENT            Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 547.
Location/Qualifiers
1..791
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/notes="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."
BASE COUNT        251 a 138 c 139 g 260 t 3 others
ORIGIN

Query Match       21.3%; Score 762.2; DB 10; Length 791;
Best Local Similarity 98.8%; Pred. No. 1.2e-144;
Matches 778; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 2038 agtgtttgaagatgtatgtgacacctccagacacagaagaaactctagcaaa 2057
|||||
Db 3 AGTGTTTGAAGATGATGGTGATCTTGAACCTCCAGACACAGAAACTCTAGCAAA 62

QY 2098 ttcaggggaagtgtgtacacactcaggctgcagtgattttcagcaaaacttgattggacaaa 2157
|||||
Db 63 TTCAGGGGATGTTGTCTACACTCAGGCTGCAGTATTTTCAGCAAACTTGATTGGACAAA 122

QY 2158 cgggcctgtcccttatcttttggtagtgaaaaaatttgagctagtggaagcgaatacgt 2217
|||||
Db 123 CGGGCCCTGTGCCTATCTTTTGTGGAGTGAAAAAATTTGAGTAGTGGAAGCCAAATCGT 182

QY 2218 aactacagaacgacgatcagcatcacctggctcttctgctgattgcaaataggcattta 2277
|||||
Db 183 AACTTACGCAAGCAGCATGCAGCATACCTTGGCTCTTGTGTTGTAATGCAATGCAATTTA 242
```

```
QY 2278 aaatggaatttgaatcagatgctcctcattctccagttaaagtggcatcagtggt 2337
|||||
Db 243 AAATGTGAATTTGAATCAGATGCTCTTTTATTTCAGTTTAAAGTGGCATCATAGGTGT 302

QY 2338 ttcttaagttttaagtcttggtataaaaaactccaccagtgctctaccatctccaccatgaac 2397
|||||
Db 303 TTCTTAAGTTTAAAGTCTCGATGATAAAAAACCTCCACCAGTGTCTACCATCTCCACCACATGAAC 362

QY 2398 tcgtttaaggaagcttcattcttttggtatattccgcgctcttttctcatttccctgcttt 2457
|||||
Db 363 TCGTGTTAAGGAAGCTTCATTTTGTATATTCCCGCTCTTTCTCTCTTATTTCCAGTCTT 422

QY 2458 ctgcataatcatgctctcttctgtaagtaattcaagcataagatcttgggaataataaaatc 2517
|||||
Db 423 CTGCATAATCATGCCCTTCTTGTCTAAGTAATTCRAAGCATAAAGATCTTGGATAATAAATAATC 482

QY 2518 acaatcttaggagaagaataaaaaattgttattttcccgctcttcttcccgctcttcccgctctt 2577
|||||
Db 483 ACAATCTTAGGAGAAAGATAAAATTTGTTATTTTCCAGTCTCTTTCCAGCATGATGATATC 542

QY 2578 ttatgattaaaaacaaatttaaaatttttaaaacacctgaagataaaattagaagaaattgtgc 2637
|||||
Db 543 TTATGATTAATAAACAAATTTAAATTTTAAACACCTGAAGATAAAATTTAGAGAAATTTGTGC 602

QY 2638 acctccacaaaacatacaaaagtttaaagtttggtatcttttctcagcaggtatcagtt 2697
|||||
Db 603 ACCCTCCACANAACATACAAAGTTTAAAGTTTGGATCTTTTCTCAGCAGGTATCAGTT 662

QY 2698 gtaataatgaattagggcccaaaatgcataaacgaaaaaagaagcagctacatagttta 2757
|||||
Db 663 GTAATAATGAATTAGGGGCCAAATGCANAAACGAAAAAATGAAGCAGCTACATGATGTTA 722

QY 2758 gtaattctagttgaaactgtaattgtaattgttggtcttcattgtatttttatattg 2817
|||||
Db 723 GTAATTTCTAGTTGAATGTAATTG-ATATTGGGGCTTCATATGTTATTTAATATTG 781

QY 2818 tactttttt 2826
|||||
Db 782 TACTTTTTT 790

RESULT 5
BG216627
LOCUS             RST36320 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
DEFINITION
ACCESSION          BG216627
VERSION            BG216627.1 GI:13742648
KEYWORDS           EST.
SOURCE             human.
ORGANISM           Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS            Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Costanzo,D., McElligott,K., Boozer,S., Mays,R., Smith
,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher
,J., Danzig,J. and Ducar,M.
TITLE              Creation of genome-wide protein expression libraries using random
activation of gene expression
JOURNAL            Nat. Biotechnol. 19 (5), 440-445 (2001)
MEDLINE
COMMENT            Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 557.
Location/Qualifiers
1..786
/organism="Homo sapiens"
/db_xref="taxon:9606"
```

<div><div>/clone_lib-"Athersys RAGE Library"</div><div>/cell_line="Hr1080"</div><div>Note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is Hr1080, since a random activation method was used, these sequence tags are not necessarily expressed in Hr1080 under normal circumstances."</div></div>									
BASE COUNT	252 a	152 c	145 g	234 t					
ORIGIN									
Query Match	20.7%	Score	741.6;	DB	10;	Length	786;		
Best Local Similarity	98.8%	Pred.	No. 1.8e-140;						
Matches	747;	Conservative	0;	Mismatches	9;	Indels	0;	Gaps	0
Qy	1949	ttatacacaaaatgaccagggccagagactctacccaaagaatgggtttgtatttaagggc	2008						
Db	30	TTATACACAAAACGACCAGGCCAGGACTCTACCCAAAGATGGTTCGTATTAAAGGCG	89						
Qy	2009	cccagcagttagaacatcctcagaaaaagaagtgtttgaaagatgctatggtgactcttga	2068						
Db	90	CCCAGCAGTTAGAGCATCCTCAGAAAAAGAGTTTGAAGATGATGGTGATCTTGAAA	149						
Qy	2069	cttcagacacaaagaaaactctagcaaatcaggggaagttgtctacactcagcctgc	2128						
Db	150	CCTCCAGACACAGAAAACCTCTACCAAAATCAGGGGAAGTTTGCTACACTCAGGCTCG	209						
Qy	2129	agtatcttcagcaaaacttgattgggacaaacgggcctgtgccttatctctttgggtggagtga	2188						
Db	210	AGTATTTTCAGCAAACTTGATTGGCAACGGGCGCTGTGCCTTATCTTTTGGTGGAGTGA	269						
Qy	2189	aaaaatttgagctagtagcaccaaatcgttaacttacagcaagcagcatgcagcatacctg	2248						
Db	270	AAAAATTTCAGCTAGTGAAGCCAAATCAATACTTACAGCAAGCAGCATGCAGCATACCTG	329						
Qy	2249	gctctttgctgattgcaaataggcattttaaagtgtgaattgggaatcagatgctctccatt	2308						
Db	330	GCTCTTTCGTGATTGCAATAGGCATTTAAATGTGAATTTGGAAATCAGATGCTCTCCATT	389						
Qy	2309	acttcagcttaaatggtgcatacataggtgttctoctaagttttaaagttcttggataaaaactc	2368						
Db	330	ACTTCCAGTTTAAAGTGGCATCATAGTGTTTCCCTAAGTTTAAAGTCTTGGATAATAACTC	449						
Qy	2369	caccagttctacatctccaccatgaactctgttaaaggaagcttcatttttgatatcttc	2428						
Db	450	CACCAGTGTCTACCATCTCCACCTGAACTTGTGTTAAGGAAGCTTCATTTTGTATATTC	509						
Qy	2429	ccgctctttctcttcatttccctgtctctcgtcataaatacatgctcttctgctaagtaatt	2488						
Db	510	CGGCTCTTTCTCTTCATCTCCCTGCTCTTCGCATAAATCATGCCCTCTTGCTAAGTAATT	569						
Qy	2489	caagcataagatcttggaaataataaaaaatacaatcttaggagaagaataaaaattgttat	2548						
Db	570	CAAGCATTAAGATCTTGGAAATAATAAAATCACAATCTTAGGAGAGAGAATAAAATTTGTTAT	629						
Qy	2549	ttcccaagctctctggccatgatgatattctatgattaaaaacaataataatttttaaac	2608						
Db	630	TTTCCAGCTCTCTTGGCCATGATGATCTTATGATTAATAAAACAAATTAATAATTTTAAAC	689						
Qy	2609	acctgaagataaaattagaagaaattgtgcaccctcccaaaaaacatacaaaagttttaaagt	2668						
Db	690	ACCAGAAGATAAATTAGAAGAAATTTGTCCACCTCCACAAACATACAAAGTTTAAAGT	749						
Qy	2669	ttgatctttttctcagcaggtatcatcagttgtaata	2704						
Db	750	TTGGATCTTTTCTCAGCAGGTATCAGTTGTAAAAA	785						

RESULT	6
LOCUS	BG193725 798 bp mRNA linear EST 21-APR-2001
DEFINITION	RST12861 Atherys RAGE Library Homo sapiens cDNA, mRNA sequence.

ACCESSION	BG193725
VERSION	BG193725.1
KEYWORDS	GI:13715412
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 798) Harrington,J.,J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R., Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.J., , Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith ,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher ,J., Danzig,J. and Ducar,M. Creation of genome-wide protein expression libraries using random activation of gene expression Nat. Biotechnol. 19 (5), 440-445 (2001)
TITLE	21227151
JOURNAL	Contact: Scott J. Cain
MEDLINE	Athersys, Inc.
COMMENT	3201 Carnegie Ave, Cleveland, OH 44115, USA Tel: 216 431 9900 Fax: 216 361 9596 Email: scain@atersys.com High quality sequence stop: 548. Location/Qualifiers 1..798 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="Athersys RAGE Library" /cell_line="HT1080" /note="See 'creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."
BASE COUNT	257 a 140 c 142 g 259 t
ORIGIN	
Query Match	19.9%; Score 715.2; DB 10; Length 798;
Best Local Similarity	97.2%; Pred. No. 4.3e-135;
Matches	770; Conservative 0; Mismatches 18; Indels 4; Gaps 4;
Qy 2043	tTgaagaatgatggtgatcttTgaacctcagacacaagaacttttagcaaatccag 2102
Db 8	TTACAAGAGATGGTGCTGTGAACCCTCAGACACAGAAGAAACTTCTACCAATTGAG 67
Qy 2103	gggaagtgttgtcacactcaggctgcagtagatttcagcaaaacttgattggcaaacggcg 2162
Db 68	GGGAAGTTTGCTACACTCAGGCTGCAGTATTTTCAGCAAACCTTGATTGGACAAACGGGC 127
Qy 2163	ctgtgccttatcttttggTggagtgaaaaatttgagctagtaggaagccaatcgtaactt 2222
Db 128	CAGTGCCTTATCTTTTGGTGGAGTGAAAAAATTTGAGCTAGTAGAAGCCAAAATCGTAAC TT 187
Qy 2223	acagcaagcagcatgcagcatcacctggctcttctgattgccaataggcatttaaagt 2282
Db 188	ACAGCAAGCAGCATGCAGCATACCTGGCTCTTCTGATTGCCAANTAGGCAATTTAAATG 247
Qy 2283	Tgaatttggaaatcagatgtctcccattacccttaaagtggcatcatagtggttttctc 2342
Db 248	TGAATTTGGAATCAGATGCTCTCCATTACTTCCAGTTAAAGTGGCAACATAGTGTCTTCCT 307
Qy 2343	aagttttaagtcttTggataaaaactccaccagtgcttaccactctccaccatgaactctgt 2402
Db 308	AAGTTTTTAAGTCTTTGGATAAAAACTCCACCAGTCTCTACCATCTCCACCATGAAC TGT 367
Qy 2403	Taaggaaagctcatttttgatatattccgcgtcttttctcttcttccattccctgtctctgca 2462
Db 368	TAAAGAGAGCTTCATTTTGTGTATATTCGGCGTCTTTTCTTCATTTCCCCTGCTCTTCGCA 427
Qy 2463	Taatcatgcctctcttgctaagtaattcaagcataagaatcttggataataataaaacaaat 2522

Db	428	TAATCATGCCCTCTTGCTAAGTAATTCACGATCAAGATCTTGGAAATAATAAAAATCACAAT	487
Qy	2523	cttagagagaagaataaaaattgttattttccagctctcttggccatgatgatattcttatg	2582
Db	488	CTTAGGAGAAAGAAATAAAAATGTGTTATTTCCCACTCTCTGGCCATGATGATATCTTATG	547
Qy	2583	attaaaaacaaaataaa-tttttaaaccactgaagataaaattagaagaaaattgtgcacc	2641
Db	548	ATTAAAAACAATTAATAATTTTTTAAACACCTGAAGATAAATTAAGAAGAAATGTGCACCC	607
Qy	2642	tcacaaaaacatacaaaagttttaaagtttgtagtctctttctcagcaggtatcagttgtaa	2701
Db	608	TCACAAAACATCAAAAGTTTAAAGTTTGGATCTTTTTCTCAGCAGGTATCAGTTGTAA	667
Qy	2702	ataatgaattcaggggccaaatgcacaaacgaaaaatgaacgagctacatgattagtaa	2761
Db	668	ATAATGAATTTAGGGGCAAAATGCCAAACG-AAAATGAACGACGTACATGTAGTTAGTAA	726
Qy	2762	-tttctagtttgaactgaattgaatatt-gtggctcatatgtattatttattatgta	2819
Db	727	TTTTCTAGGTTGAACGTAAATGAATATTGGGGCTTCATATGAATTAATTATTATATCGGA	786
Qy	2820	cttttttcattta	2831
Db	787	CTTTTTTTATTTA	798

RESULT		7	
BG209741			
LOCUS	RST23267	Athersys RAGE Library Homo sapiens cDNA,	mRNA linear EST 21-APR-2001
DEFINITION			
ACCESSION	BG209741		
VERSION	BG209741.1	GI:13731428	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
REFERENCE	1 (bases 1 to 768)		
AUTHORS	Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R., Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J. and Ducar,M.		
TITLE	Creation of genome-wide protein expression libraries using random activation of gene expression		
JOURNAL	Nat. Biotechnol. 19 (5), 440-445	(2001)	
MEDLINE	21227151		
COMMENT	Contact:	Scott J. Cain	

```

FEATURES
source
nrgn_query_sequence_group: 401.
Location/Qualifiers
1. .768
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."
247 a 151 c 229 t
BASE COUNT

```

Query Match  
18.6%; Score 666; DB 10; Length 768;

Best Local Similarity		96.0%	Pred. No. 4.3e-125;	Matches		737;	Conservative	0;	Mismatches	25;	Indels	6;	Gaps	5;
QY	1940	aggatgacttatacacacaaatgaccaggccaggagctctacccaagaatgggtttgta	1999											
Db	1	AGANTGGACTTTATACACAAATGACCGAGCCAGGACTCTACCCAAAGAATGGGTTTGTGA	60											
QY	2000	tttaaggccccagcagtgtagaacatcctcagaaaaagaagtggttgaagaatgatgggtg	2059											
Db	61	TTTAAAGGCCCCAGCAGTTAGAACATCCTCAGAAAAGAAGTGTTTGAAAGATGATGATGGTG	120											
QY	2060	atcttgaacctccagacacaaagaataactctagcaaatcaggggaagtttgtctacac	2119											
Db	121	ATCTTGAACACTCCAGACACAAAGAAATTTCTAGCAAAATTCAGGGGAAGTTGTCTACAC	180											
QY	2120	tcaggctgcagtatcttcagcaaaacttgatggacaacaggcgctgtgccttatctcttg	2179											
Db	181	TCAGGCTGCAGATTTTCAGCAAACTTGATTTGGACAACAGGGCCTGTGCGCTTATCTTTTG	240											
QY	2180	gtggagtgaaaaaatttgactagtgagcgcacaaatctaaacttacagcaagcagcatgca	2239											
Db	241	GTGGAGTGAATAATTTGAGCTAGTAGAAGCAAAATCGTAACCTTACAGCAAGCAGCATGCA	300											
QY	2240	gcatacctggctctttgctgatgcaaataggcaattttaaattggaatttggaatcagat	2299											
Db	301	GCATPACCTGGCTCTTCTGCTGATTGCAATAGGCAATTTAAATGTGAATTTGGAATCAGAT	360											
QY	2300	gtctccattactccagttaaagtgccatcatagtggtttcttccaaagttttaaagttcttgg	2359											
Db	361	GTCTCCATTACTTCCAGTTAAAGTGGCATCATAGTGTTCCTTAAGTTTAAAGTCTTTGGA	420											
QY	2360	taaaactccaccagtgctaccatctccaccatgaaactctgtaaagaagcttcattt	2419											
Db	421	TAAAAACTCCACCAGTGCTTACCATCTCCACCATGAACCTCTGTTAAAGGAAGCTTTCATTTT	480											
QY	2420	tgtatatcccgctctttctctcttcttccatcttccctgtctcttgcataatca-tgccttcttg	2478											
Db	481	TGTTATATCCCGCTCTTTTCTCTTCATTTCCCTGCTCTTGCATTAATCATTTGCCCTCTTG	540											
QY	2479	ctaagtaaticcaagcaataagatc-ttggataataataacacacatcttagagagaagaat	2537											
Db	541	CTAAGTAATTCACGATTAAGATCTTTGGATATTAATAATACATCTCTTAGGAGAAAGAT	600											
QY	2538	aaaattgtattttcccgagtcctctggccatgatgatatttatgataaaacaaatta	2597											
Db	601	AAAATTGTTATTTCCCGAGTCTTTTGGCCCTGATGATATCTTATGATTAATAAACAGATTA	660											
QY	2598	aa-ttttaaaacacctgaagataaattagaagaaa-ttgtcacccctccacaaacatac	2655											
Db	661	AATTTTAAAAACACTGGAAACATAAATAAGAACAAATTTGGCACCCCTCCACCAACATAC	720											
QY	2656	aaagttaaagaagt-ttgatctttttcttcagcaggtatcagttgtaa	2701											
Db	721	AAAGTGTAAAAAGTCCGGATCCCTTTTCTCAACAGGAATCAGATGGAA	768											
RESULT 8														
LOCUS		BC286508	736 bp	mRNA	linear	EST 21-FEB-2001								
DEFINITION		602382504F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4500121 5',												
ACCESSION		BC286508												
VERSION		BG286508												
KEYWORDS		BG286508.1 GI:13039446												
SOURCE		EST.												
ORGANISM		human.												
REFERENCE		Homo sapiens												
AUTHORS		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;												
TITLE		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.												
JOURNAL		1 (bases 1 to 736)												
COMMENT		NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .												
		National Institutes of Health, Mammalian Gene Collection (MGC)												
		Unpublished (1999)												
		Contact: Robert Strausberg, Ph.D.												



Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10365 row: g column: 02  
High quality sequence stop: 730.  
Location/Qualifiers  
1. .736  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4500121"  
/clone\_lib="NIH\_MGC\_93"  
/tissue\_type="transitional cell papilloma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/notes="Organ: bladder; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC Library."  
202 a 189 c 159 g 186 t

BASE COUNT  
ORIGIN

Query Match 18.2%; Score 653.2; DB 10; Length 736;  
Best Local Similarity 97.4%; Pred. No. 1.7e-122;  
Matches 721; Conservative 0; Mismatches 3; Indels 16; Gaps 5;  
QY 1518 atatgttgagctcttcagtgatgagatgaactttgttgaatgggggcac 1577  
Db 1 ATATGTTGAAGTCTTCAG-GTTCAGCTGAGGATGAACGTTGTATATGGGGGCAC 59  
QY 1578 tttaaatcgaatggcttatcccccagc-----ccatgcctgtctctccctc 1625  
Db 60 TTTAAATCGAAATGGCTTATCCCCACGCCCATGTAAGTTACCATGCGTCTCTCCCTC 119  
QY 1626 ctacacattccagctcctcgtcagttatctctacagaagctgcattaccagcctc 1685  
Db 120 CTACACATTTCCAGCTCCTGCTGCGAGTTATTCCTACAGAAGCTGCCATTTACCGCCCTC 179  
QY 1686 tgtgatttgaatccagcagcaactgcagcctccacagcgtactaccagcagcactca 1745  
Db 180 TGTGATTTTGAATCCACGAGCACTGCAGCCCTCCACAGCTACTACCCAGCAGCACTCA 239  
QY 1746 gctcttcataactacacagcgtactatccacgccccccaggttcgcctaatgcttgg 1805  
Db 240 GCTCTTCATGAATACACAGCGTACTATCCAGCCCCCAGGTTCCGCTAATAGTCTTGG 299  
QY 1806 ctacttccctacagctgaatcttagcgtgtccctccacagcctgcacggtggtcag 1865  
Db 300 CTACTTCCCTACAGCTGCTAATCTTAGCGGTGTCCTCCACAGCCTGGCAGGTTGGTCAG 359  
QY 1866 aatgcaggcgtgcctacataactgagtgtaaggaatcttcaacttctccaagtta 1925  
Db 360 AATGCAGGCGCTGSCCTACAACTACTGAGTTAAGGAATCTTAACTTCTTCCAGGTTA 419  
QY 1926 ccagtatgcacccagagatgaacttatcacacacaaatgaccagccaggactctaccacaa 1985  
Db 420 CCAGTATGCACCCAGAGATGGACTTATACACACAAATGACGAGCCAGGACTCTTACCCAA 479  
QY 1986 agaatgggttgatttaagggcccccagcagtttagaatacctcagaaagagtggttg 2045  
Db 480 AGAATGGG-TGTGATTTAAGGGGCCCCAGCAGTATAGAACATCTCTCAGAAAAAGAGTGTG 538  
QY 2046 aaagatgatggtgatcttgaacctccagacacaaagaaactctcagcaaatcagggg 2105  
Db 539 AAAGATGATGGTGATCTTGAACCTCCACACACAAAGAAATCTTAGCAAAATTCAGGGG 598  
QY 2106 aagttgtctacactcaggtgcagtagtatttcagcaactgatttgacaaagggcgtg 2165  
|||||

Db 599 AAGTTTCTCTACACTCAGGCTGCAGTAGTATTTTCAGCAAACTGTATGGACAAACGGGCGT 658  
QY 2166 tgcctattcttggtagtgagtaaaattgaactgaagcgaacgaatcgaacttaca 2225  
Db 659 TGCCTTATCTTTTGGTGGAGTGAAAAAATTTGACCCAGTGAAG-CAATCGTAACCTTACA 717  
QY 2226 gcaagcagcatgcagcatc 2245  
Db 718 G-AAGCAGCATGCAGAAATAC 736  
RESULT 9  
LOCUS BG217686 782 bp mRNA linear EST 21-APR-2001  
DEFINITION RST37400 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.  
ACCESSION BG217686  
VERSION BG217686.1 GI:13743707  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 782)  
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,  
Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.,  
Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith  
,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher  
,J., Danzig,J. and Ducar,M.  
TITLE Creation of genome-wide protein expression libraries using random  
activation of gene expression  
JOURNAL Nat. Biotechnol. 19 (5), 440-445 (2001)  
MEDLINE 21227151  
COMMENT Contact: Scott J. Cain  
Athersys, Inc.  
3201 Carnegie Ave, Cleveland, OH 44115, USA  
Tel: 216 431 9900  
Fax: 216 361 9596  
Email: scain@athersys.com  
High quality sequence stop: 473.  
FEATURES  
source  
1. .782  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Athersys RAGE Library"  
/cell\_line="HT1080"  
/notes="See 'Creation of Genome-wide Protein Expression'  
Libraries using Random Activation of Gene Expression',  
Nature Biotechnology, in press. Note that even though the  
cell type indicated is HT1080, since a random activation  
method was used, these sequence tags are not necessarily  
expressed in HT1080 under normal circumstances."  
BASE COUNT 243 a 130 c 144 g 245 t  
ORIGIN  
Query Match 17.8%; Score 637; DB 10; Length 782;  
Best Local Similarity 95.5%; Pred. No. 3.3e-119;  
Matches 677; Conservative 0; Mismatches 30; Indels 2; Gaps 2;  
QY 2038 agtgttgaagatgtagtgatcttgaacacctccagacacagaaactctagcaaa 2097  
Db 2 AGTCTTTGAAAGATGTATGGTGATCTTGAACACCTCCAGACACAGAAACTTCTAGCAAA 61  
QY 2098 ttcagggaagtgtgtctacactcagctgagctattttcagcaacttgattgacaaa 2157  
Db 62 TTCAGGGGAAGTTGTCTACACTCAGGCTGCAGTATTTTTCAGCAACTTGTATTGGACAA 121  
QY 2158 cggcctgtgccttatctctttgtggagtgaaaaatttgagctagtgaaagccaaatcgt 2217  
Db 122 CGGCGCTGTGCCCTATCTTTGGTGGAGTGAAAAAATTTGAGTTAGTGAAGCCAAATCGT 181  
QY 2218 aacttacagaacgagcatgcagcatcaccttgctcttctgattgcaaatagcattta 2277  
|||||



```

VERSION BI557167.1 GI:15444468
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 913)
REFERENCE NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cga@rs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1738 row: p column: 09
High quality sequence stop: 883.
FEATURES
source
1..913
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="NIH_CGAP_Mam3"
/tissue_type="tumor, gross tissue"
/lab_host="DH10B"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert 2 kb. Library constructed by Life
Technologies, catalog #12017-018. Investigators providing
samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference
for transgenic model: Xu et al., Nature Genetics 22, 37-43
(1999). Note: this is a NCI_CGAP Library."
BASE COUNT 242 a 221 c 240 g 210 t
ORIGIN
Query Match 17.6%; Score 629.6; DB 10; Length 913;
Best Local Similarity 85.8%; Pred. No. 1.1e-117;
Matches 757; Conservative 0; Mismatches 119; Indels 6; Gaps 5;
QY 371 atctctttttgttcgaaagaattcaagaattgttcctcctggcttcacctgattg 430
Db 1 ATTCCTTTTTCGATCTTCGGAAGAAGTTCAAGAAGTCTGCGGGGCTCTCCCGATATCG 60
QY 431 acaactggacgttgcacaaatgacagagtatttaatttttgagaagtagttcagttc 490
Db 61 ACAAGCTGGATGTCGCAGCGATGACAGAAAGCCTGAACTTTGAGAAGAGTGACTCTGTG 120
QY 491 ctgatatgagcctctcaagttgaagatattgggaataataatttttagcaatgatttcag 550
Db 121 CCCGATACGAGGCTCCCAAGTTGAAGACATGGGGAACATAATTTTAGCAATGATCTCCG 180
QY 551 agcttataatcacagtttttcagatccagagagtggaattacaagtttgaagtggaa 610
Db 181 AGCCTACATACAGATTTTCAGATCCAGAGAGTAACTACAAGTTTGAAGATGGCA 240
QY 611 ctgacagaagatggaacttattgatataacacccgtagtcagggcagaggtttaccat 670
Db 241 CTTGCAGCAAGATGGAATTTATTGATGACAGCAGAGTGGTGAGAGCTCGAGGTTGCGCCT 300
QY 671 gcaagtcttcagatcaagatttcgaagattcttcaaaagactcaaatattgccaaaggag 730
Db 301 GGCAATCTTCAGATCAAGATTTTCGAAGGTTCTTCAAGGACTCAATATTGCCAAGGGTG 360
QY 731 gtccagcactttctgaatgctcagggtcgaaggacgagagctcgttaggtttg 790
Db 361 GCGAGCAGCATTTCCTGAATGCTCAGGGCCGGAGGAACGAGGAGCTCTGGTAGATTG 420
QY 791 taagtggaggacccgagacctagcactacagagggcacaacaacatcacatggggaccggt 850
```

```

Db 421 TAAGTGAGGACACAGAGATCTAGCCCTGCAGAGGCACAAACATCACATGGGGACCCGGT 480
QY 851 atattgaggtttcaaaagcaacaggtgaagatttccttaaaattgctggtggttactcca 910
Db 481 ACATTGAGGTTTACAAAGCAACAGCGGAAGATTTTCCTTAAATTTGCCGGAGGGACATCCA 540
QY 911 ataggttagccagttctctccaagaaaaatcaagtcattgttcgcagtcgaggggtcc 970
Db 541 ATGAGGTAGCCCAAGTTTCTCTCCAAAGGAAAAACCAAGTCATCGTGGCGCATGAGGGGGCTCC 600
QY 971 ctttcacggccacagctgaagaagtggtgcctcttttgacagcattgcccctattactg 1030
Db 601 CTTTCACGGCCACGCGTGAAGAAGTGTGCGCTTCTTTTGGACAGCACTGCCCATCACTG 660
QY 1031 ggggaaaggaagcattcct-ctttgtcacctaccacagatggttaggcca-agggggagcg 1088
Db 661 GAGGGAAGGAAGCATCTCTCTTCTGTTACCTACCCAGATGTAGGCCAACCCGGAGATGC 720
QY 1089 ttgtgctctttgctgtgaggaatgacagaatgcgctgaggaagcagataaagactt 1148
Db 721 TTTTGTGCTCTTGTGCTGTGAGGAGTATGCTCAGAACGCCCTGAGGAAGCACAAGGGAGTT 780
QY 1149 gttgggttaaaagatacatt-gaactcttcaggagcagcagctggaagttcagcaggtgc 1207
Db 781 GCTGGGTAAAAGATACCTTCGAGCTCTTCAGGAGCAGAGTGGCGAAGTCAGCAAG--TG 838
QY 1208 tgaatgattctcctcggccctctcattccacttcaaaccc 1249
Db 839 TTAACCGATTCTCCTCAG-CCCTCTGAATCCACTTCCAACCC 879
RESULT 12
BG287429 960 bp mRNA linear EST 21-FEB-2001
LOCUS 602382018F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4499644 5',
DEFINITION mRNA sequence.
ACCESSION BG287429.1 GI:13041262
VERSION BG287429
KEYWORDS ESI.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 960)
REFERENCE NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cga@rs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10364 row: c column: 05
High quality sequence stop: 632.
FEATURES
source
1..960
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_93"
/tissue_type="transitional cell papilloma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: bladder; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 326 a 183 c 151 g 299 t 1 others
```

ORIGIN

Query Match 17.5%; Score 627.6; DB 10; Length 960;  
 Best Local Similarity 99.2%; Pred. No. 2.7e-117;  
 Matches 641; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 2366 ctccaccagtgctaccatctccaccatgaactcgtttaagaagctcttcattttgtata 2425  
 Db 1 CTCCACCAGTGCTACCATCTCCACCATGAACCTGTTAAGGAAGCTTCATTTTCGTATA 60

QY 2426 ttccgcgtttttctctcatttccctgtcttctcacaataatcctcttctgtcaagta 2485  
 Db 61 TTCCCGCTTTTCTCTCAATTTCCCTGCTCTTGCATAATCATGCTCTTCTGCTAAAGTA 120

QY 2486 attcaagcataagattcttggaataataacacatctcttagagagaagaataaaattgt 2545  
 Db 121 ATTCAGCATATAGATCTTGGATATATAAATCACAATCTTAGAGAGAAGAAATAAATGT 180

QY 2546 tattttcccgctcttggccatgatgatcttcttgatgataaaacaaattaaatttaa 2605  
 Db 181 TATTTTCCCGCTCTTGGCCATGATGATATCTATGATTAAACAAATTAATTTTAA 240

QY 2606 acacacctgaagataaattagaagaattgtgacccctccacaaacatacaaaagtttaa 2665  
 Db 241 AACACCTGAAGATATATTAGAGAAATTTGCACCCCTCCACAAACATACAAAGTTTAA 300

QY 2666 agtttgatcttttctcagcaggtatcagttgtaataataatagaggccaaaatgc 2725  
 Db 301 AGTTGGATCTTTTCTCAGCAGGTATCAGTTGTAATAATGAATAGGGGCCAATGC 360

QY 2726 aaacgaaataaagaagcagctacatgtagttagtaattctgtttgaactgtaattgaa 2785  
 Db 361 AAAACGAAATCAAGCAGCTACATGCTAGTATGTAATTTCTAGTTGAACTGTAATTGAA 420

QY 2786 tatttggttccatgattattttattatttatttatttatttatttatttatttatttggac 2845  
 Db 421 TATTGTTGGCTTCATATGATATATTTTATATTGTATGTTTCTTATTAATGATGTTGGAC 480

QY 2846 tttaataagagaaattccctagtttttataatcccgagagtgagacaatttgaacagtg 2905  
 Db 481 TTTAATAGAGAAATTCATAGTTTATTAATATCCAGAGAGTGAGACAAATTTGAACAGTG 540

QY 2906 attctgaaacaaatacactaactgaacgaagtgatgtaattcttatatatattatgatagcc 2965  
 Db 541 ATTCTAGAAAACAAATACACTAACTGAACAGAGTGAATGCTTATATATATTATGATAGCC 600

QY 2966 ttaaaccttttctccttaagccttaactgaactgaataataataaacc 3011  
 Db 601 TTAACCC-TTTTCTCTAATGCTTAACTGTCANATTAATTAATTAAC 645

RESULT 13  
 BG184305 819 bp mRNA linear EST 21-APR-2001  
 LOCUS RST3229 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.  
 DEFINITION BG184305  
 ACCESSION BG184305  
 VERSION BG184305.1 GI:13705992  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 819)  
 AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,  
 Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.,  
 Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith,  
 J., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher,  
 J., Danzig,J. and Ducar,M.  
 TITLE Creation of genome-wide protein expression libraries using random  
 activation of gene expression  
 JOURNAL Nat. Biotechnol. 19 (5), 440-445 (2001)  
 MEDLINE 2127151

COMMENT

Contact: Scott J. Cain  
 Athersys, Inc.  
 3201 Carnegie Ave, Cleveland, OH 44115, USA  
 Tel: 216 431 9900  
 Fax: 216 361 9596  
 Email: scain@atersys.com  
 High quality sequence stop: 382.

FEATURES

source  
 1..819  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Athersys RAGE Library"  
 /cell\_line="HT1080"  
 /note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT 250 a 177 c 152 g 240 t  
 ORIGIN

Query Match 17.3%; Score 619.8; DB 10; Length 819;  
 Best Local Similarity 90.4%; Pred. No. 1e-115;  
 Matches 742; Conservative 0; Mismatches 67; Indels 12; Gaps 7;

QY 1928 agtatgaaccagagatggacttatcacacaaatgaccaggccagagctaccaccaag 1987  
 Db 1 AGTATGCAACCCAGGATGGACTTATACACACAAATGACCAGGCCAGGACTCTACCCAAAG 60

QY 1988 aatgggtttgtatttaagggcccgagcagttagagaacatctcagaaaaaagttgttaa 2047  
 Db 61 AATGGGTTTGTATTTAGGGGCCCGACAGTTAGAACATCTCAGAAAAAGAAAGTTTGAA 120

QY 2048 agatgtatgtgatcttgaacctccagacacaaagaacttctagcaaatccaggggaa 2107  
 Db 121 AGATGTATGGTGATCTTTGAACCTCCAGACACAGAAAACTTCCAGCAATTCAGGGGAA 180

QY 2108 gttgtctacactcaggtgcagtagtatttcagcaaaacttgaggcaaacgggctgtg 2167  
 Db 181 GTTTGCTACACTCAGGCTGCAGTATTTTCAGCAAACTTGATTGGACAAACGGGCTGTG 240

QY 2168 ccttatcttttggtagagtgaaaaatttgagctagtgaagccaaatcgaacttcacgc 2227  
 Db 241 CCTTATCTTTTGGTGAGTGAAAAATTTGAGCTAGTGAAGCCAAATCGTAACCTTACGC 300

QY 2228 aagcagcatgcagcatcacctggctctttgctgattgcaaatagcatttaaaatgtgaat 2287  
 Db 301 AAGCAGCATGCAGCATACCTGGCTCTTTGCTGATTGCAAAATAGGCATTTAAATGTGGAT 360

QY 2288 ttggaaatcagatgtctccattacttccagttaaagtggcagtcaggtgttctctaagtt 2347  
 Db 361 TTGGAATCAGATGCTCCCATTACTTCCAGTTAAAGTGGCATCATAGTGTGCTTAAGTT 420

QY 2348 ttaagcttggataaaactccacagctcttaccatctccaccatgaactctgttaagg 2407  
 Db 421 TTAAGCTTTGGATAAAAACTCCACCAGTGTCTACCATCTCCACCATGAATCTGTTAAGG 480

QY 2408 aagcttcattttgtatattcccgctctttctctcaatttccctgtcttctgcataatc 2467  
 Db 481 AGCTTTCATTTTGTATATATCCCGCTCTTCTCTCTTTTCATTTCCCTGCTCTCTTAATC 540

QY 2468 atgcc-ttcttgcagtaagtaattcaagcataagatctcttggaataataaataacactta 2526  
 Db 541 ATGCCCTTCTTGTGAAGCCACCACCAAGCTTAAGATCTTGGCATATTCAAATC-CCTTCTTA 599

QY 2527 ggagaagaataaaattgttttttccagctctcttggccatgatgatctcttatgatta 2586  
 Db 600 GGAGAACG-ATAAAATGGTGTCTTTCCAGTCTCTTGGCCCTGTATGATCTCTTACGAAAT 658

QY 2587 aaacaaaattaaattttaaacacc-tgaagataaaattagaagaaattgt-gcaccctcc 2644

```
Db 659 AAAACATACTAACTTTAAACCCCTTGAAGAAAAATTAGAAGAAATTTTGGCCCCCTTC 718
Qy 2645 acaaacatacaagtttaaaagtttg---atcttttctcagcaggtatcagttgtaa 2701
Db 719 CCAACATACAAAGGCTAAACGTTTGATCTCTTTTTCAGCAGGTTTCCCTGTTTA 778
Qy 2702 ----ataatgaattagggcccaaatgcaaacagaaaaatg 2738
Db 779 CATTAATGAATATAGGGCCCAAAATAGAAAACCGATAAG 819

RESULT 14
BE870387
LOCUS BE870387 904 bp mRNA linear EST 20-OCT-2000
DEFINITION 60144757f1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3851705 5',
mRNA sequence.
ACCESSION BE870387
VERSION BE870387.1 GI:10319163
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 904)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM9572 row: m column: 18
High quality sequence stop: 622.
Location/Qualifiers
1..904
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3851705"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 283 a 142 c 188 g 291 t
ORIGIN

Query Match 17.2%; Score 618; DB 10; Length 904;
Best Local Similarity 95.1%; Pred. No. 2.4e-115;
Matches 702; Conservative 0; Mismatches 30; Indels 6; Gaps 6;

Qy 2753 agttagtaattcttagttgaaactgaattgaattatgtgcttcattgtattttta 2812
Db 1 AGTTAGTAATTTCTAGTTTGAACGTAAATGAATATTTGGCTTCATATGATTATTTTA 60
Qy 2813 tatgtactttttcattatgatgttgcttgaactttaagaagaaattccatagttttt 2872
Db 61 TA-TGTACTTTTTCATTTATTGATGGTTTGACCTTTAATAAGAGAAATCCATAGTTT 119
Qy 2873 aatatccagaaagtgaacaatttgaacagtgtattcttagaaaaaacatacaactgaa 2932
Db 120 AATATCCCAAGAGTGAGACAATTTGAACAGTGATTCTTAGAAAACAATACACTAACTGAA 179
Qy 2933 cagaagtgaatgcttatatatattatgatagccttaaacctttttcctcctaagccttaa 2992
Db 180 CAGAAGTGAATGCTTATATATATTATGATAGCCTTAAACCTTTTTCCTCTAATGCTTAA 239
```

```
Qy 2993 ctgtcaataaattaaacctttttaagcataggactatagtcagcatgctagactgagag 3052
Db 240 CTGTCAAAATATTATAACCTTTTAAAGCATAGGACTATAGTCAGCATGCTAGACTGAGAG 299
Qy 3053 gtaaacactgatgcaattagaaacaggtactgatgctgcagtggtttaacactatgttag 3112
Db 300 GTAAACACTGATGCAATTTAGAACACAGGTACTGATGCTGTGTCAG-GTTTAAACACTATGTTTAG 358
Qy 3113 ctgtgttatgctataaaagtcaaatattagacacactagctagctactgctcctcatgtaa 3172
Db 359 CTGTGTTTATGCTATAAAAGTGCAATATTAGACACTAGTAGTACTGCTGCTCATGTAA 418
Qy 3173 ctccaaagaaacagagatttcattaaagtgcattgaatggatatttctctaaagttaact 3232
Db 419 CTCCAAAGAAACAGGATTTCATTAAAGTGCA-TGAATGGGATATTTCTCTAAGTTACTC 477
Qy 3233 atattgctctttgcttgaatgcaatgccgtgcagatttatgaggctgctattttatttt 3292
Db 478 ATA-TGTCCTTTGCTTGAATGCAATGCCGTGCAGATTTATGAGGCTGCTATCTTTATTTT 536
Qy 3293 ctgtgcattactttaaacaccttaagggagagaagcaaacatttccctcttca-gctgactg 3351
Db 537 CTGTGCATTACTTTAACACACCTTAAAGGGAGAGAAACATATFCCCTTCTCAGGCTGACTG 596
Qy 3352 gcaatggcccttttaactgcaataggaagaaaaaaaggtttgtgtaaaattggtga 3411
Db 597 GCAATGGCCCTTTAACTGCAATAGGAACGAACACACAGGCTCTGTGTGAACATGGGTG 656
Qy 3412 taactggcacttaagatcgaaaagaattctgtatactattgcttaagatgcttaagatgccccaa 3471
Db 657 ATAATGCACCTTTTAAGATCGACAAG-AGTTTCTGTATTGATGCTTGAAGATGCAAAAG 715
Qy 3472 gctgcccaaacgctctgaa 3489
Db 716 GTGGCCCAAGGTTCTGAA 733

RESULT 15
BG260588
LOCUS BG260588 890 bp mRNA linear EST 13-FEB-2001
DEFINITION 602372144f1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4480163 5',
mRNA sequence.
ACCESSION BG260588
VERSION BG260588.1 GI:12770404
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 890)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM10313 row: g column: 12
High quality sequence stop: 654.
Location/Qualifiers
1..890
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4480163"
/clone_lib="NIH_MGC_93"
/tissue_type="transitional cell papilloma, cell line"
/lab_host="DH10B (phage-resistant)"
FEATURES
source
```

/note="Organ: bladder; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 1.7 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."

BASE COUNT		306 a	125 c	173 g	286 t
ORIGIN					
Query Match		17.2%	Score 615.8;	DB 10;	Length 890;
Best Local Similarity		97.58;	Pred. No. 6.7e-115;		
Matches 668;		Conservative 0;	Mismatches 12;	Indels 5;	Gaps 4;
QY	2571	tgatatcttattgattaaacacaaatttaaaacacccctgaagataaattagaagaa	2630		
Db	1	TGATATCTTATGATTAAAAACAATTAATTTAAAAACACCTGAAGATATATTAGAAGAA	60		
QY	2631	atttgacacccctcacaaacacatacaaaagtttaaaagtctttttctcagcaggt	2690		
Db	61	A-TGTGCACCCCTCCACAAAACATACAAAGTTTAAAGTTTGGATCTTTTCTCAGCAGGT	119		
QY	2691	atcagttgtaataatgaattaggcgccaaatgcacaaacgaaaaatgaagcagctacat	2750		
Db	120	ATCAGTTGTAATGAATAGGCGGCCAAATGCAAAACGAAAAATGAAGCAGCTACAT	179		
QY	2751	gtagttagtaattctctagtttgaactgtaattgaattatgtgtgcttcattgtattatt	2810		
Db	180	GTAGTTAGTAATTTCTAGTTGAACCTGAATTTGAATATTGTGGCTTCATATGATTATTT	239		
QY	2811	tatatgtactttttcattattgatgttggactttaagaagaaattccatagttt	2870		
Db	240	TATATTGTACTTTTTCATTATTGATGG-TTGGACTTTTAATAAGAGAAATTCCTATAGTTT	298		
QY	2871	ttaatatcccagagtgagacaatttgaacagtgattcttagaaaaacaatacacactaactg	2930		
Db	299	TTAATATCCCAGAGTGAGACAATTTCAACAGTGTATTCTAGAAAAACAATACACTAACTG	358		
QY	2931	aacagaagtgatgcttatatatattatgatagccttaaaccttttccctcctaagcctt	2990		
Db	359	AACAGAAGTGAATGCTTATATATATATATATAGTACGCTTTAAACCTTTTCCCTAATGCGCTT	418		
QY	2991	aactgtcaataattataaaccttttaagcataggaactatagtcagcatgctagactgag	3050		
Db	419	AACTGTCAATAATATAACCTTTTAAGCATAGGACTATAGTCAGCATGCTAGACTGAG	478		
QY	3051	aggtaaacactgatgcaattagaacaggtactgatgctgcagtggtttaacactatgttt	3110		
Db	479	AGGTAAACACTGATGCAATTAGAACAGGTACTGATGCTGTCAGTGTTTAACACTATGTTT	538		
QY	3111	agctgtgttatgctataaaagtycaataattagacactagctagctgctgctcatgt	3170		
Db	539	AGCTGTGTTTATGCTATAAAGTGCATATTAGACACTAGCTAGTACTGGTG-CTCATGT	597		
QY	3171	aactccaagaacacagagatttcattaaagtcattgaatggaatttctctaaagttac	3230		
Db	598	ACTCCCAA--GAAACCGGATTTCAATTAAGTGCATTTGAATGTGGATATTCTCTAAGTTAC	655		
QY	3231	tcattatgctctttgcttgatgca	3255		
Db	656	TCATATTGGTCCCTTGGTTGAATGCA	680		

Search completed: August 15, 2002, 08:56:29  
Job time: 15240 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2002, 06:04:49 ; Search time 383.11 seconds  
(without alignments)  
2298.546 Million cell updates/sec

Title: US-09-697-206A-1

Perfect score: 3585

Sequence: 1 cttttgggacactgctgg.....gctgtttaactctcaaaaaa 3585

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCIRUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
c 1	305	8.5	351	4	US-09-328-111-380	Sequence 380, App
c 2	55.8	1.6	19124	2	US-08-487-826B-13	Sequence 13, Appl
c 3	49.8	1.4	7218	1	US-08-232-463-14	Sequence 14, Appl
c 4	47.8	1.3	602	1	US-08-764-100-8	Sequence 8, Appl
c 5	47.8	1.3	643	1	US-08-764-100-7	Sequence 7, Appl
c 6	47.8	1.3	2993	1	US-08-764-100-2	Sequence 2, Appl
c 7	47.8	1.3	2993	1	US-08-764-100-2	Sequence 10, Appl
c 8	47.8	1.3	3001	1	US-08-764-100-1	Sequence 1, Appl
c 9	44.6	1.2	837	4	US-08-998-416-288	Sequence 288, App
c 10	43.8	1.2	636	4	US-08-998-416-1137	Sequence 1137, App
c 11	43.8	1.2	642	1	US-08-764-100-13	Sequence 13, Appl
c 12	43.8	1.2	3000	1	US-08-764-100-9	Sequence 9, Appl
c 13	43.8	1.2	5852	1	US-07-867-106-2	Sequence 2, Appl
c 14	43.8	1.2	7218	1	US-08-232-463-14	Sequence 14, Appl
c 15	41.6	1.2	615	4	US-08-998-416-186	Sequence 186, App
c 16	41.6	1.2	1431	2	US-09-316-083-2	Sequence 2, Appl
c 17	41.4	1.2	19124	2	US-08-487-826B-13	Sequence 13, Appl
c 18	40.8	1.1	10968	2	US-08-680-327-2	Sequence 2, Appl
c 19	40.8	1.1	10968	4	US-03-228-246-1	Sequence 1, Appl
c 20	40.2	1.1	1939	1	US-07-715-751B-2	Sequence 2, Appl
c 21	40.2	1.1	5506	4	US-09-004-838-93	Sequence 93, Appl
c 22	40	1.1	51952	3	US-08-947-823-1	Sequence 1, Appl
c 23	39.8	1.1	3095	6	5231168-1	Patent No. 5231168
c 24	39.8	1.1	5852	1	US-07-867-106-2	Sequence 2, Appl
c 25	39.6	1.1	636	4	US-08-998-416-1137	Sequence 1137, App
c 26	39.6	1.1	731	1	US-08-451-405A-2	Sequence 2, Appl
c 27	39.6	1.1	827	4	US-08-998-416-535	Sequence 535, App

c 28	39.4	1.1	665	4	US-08-998-416-937	Sequence 937, App
c 29	39.2	1.1	6243	2	US-09-056-075-1	Sequence 1, Appl
c 30	39	1.1	1236	6	520236-4	Patent No. 520236
c 31	39	1.1	5630	2	US-08-937-931-1	Sequence 1, Appl
c 32	39	1.1	5630	4	US-09-285-502-1	Sequence 1, Appl
c 33	39	1.1	5630	4	US-09-709-126-1	Sequence 1, Appl
c 34	38.6	1.1	615	4	US-08-998-416-186	Sequence 186, App
c 35	38.6	1.1	6124	4	US-08-213-419B-3	Sequence 3, Appl
c 36	38.6	1.1	6243	2	US-09-056-075-1	Sequence 1, Appl
c 37	38.2	1.1	15766	4	US-09-338-907-73	Sequence 73, Appl
c 38	38.2	1.1	15766	4	US-09-218-207-73	Sequence 73, Appl
c 39	38.2	1.1	37950	4	US-09-338-907-183	Sequence 183, App
c 40	38.2	1.1	37950	4	US-09-218-207-183	Sequence 183, App
c 41	38	1.1	767	4	US-08-998-416-472	Sequence 472, App
c 42	38	1.1	782	4	US-08-998-416-224	Sequence 224, App
c 43	37.8	1.1	662	4	US-08-998-416-185	Sequence 185, App
c 44	37.8	1.1	701	4	US-08-998-416-701	Sequence 701, App
c 45	37.8	1.1	724	4	US-08-998-416-683	Sequence 683, App

#### ALIGNMENTS

RESULT 1  
US-09-328-111-380/c  
; Sequence 380, Application US/09328111  
; Patent No. 6262333  
; GENERAL INFORMATION:  
; APPLICANT: Endege, Wilson O.  
; APPLICANT: Steinmann, Kathleen E.  
; APPLICANT: Astle, Jon H.  
; APPLICANT: Burgess, Christopher C.  
; APPLICANT: Bushnell, Steven E.  
; APPLICANT: Carroll III, Eddie  
; APPLICANT: Catino, Theodore J.  
; APPLICANT: Derti, Adnan  
; APPLICANT: Ford, Donna M.  
; APPLICANT: Lewis, Marcia E.  
; APPLICANT: Monahan, John E.  
; APPLICANT: Schlegel, Robert  
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION  
; FILE REFERENCE: CCD-257 (US)  
; CURRENT APPLICATION NUMBER: US/09/328,111  
; CURRENT FILING DATE: 1999-06-08  
; EARLIER APPLICATION NUMBER: US 60/088,801  
; EARLIER FILING DATE: 1998-06-10  
; NUMBER OF SEQ ID NOS: 850  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 380  
; LENGTH: 351  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-328-111-380

Query Match 8.5%; Score 305; DB 4; Length 351;  
Best Local Similarity 96.3%; Pred. No. 5, 2e-71;  
Matches 338; Conservative 0; Mismatches 0; Indels 13; Gaps 2;

Qy	1388	gtactctgggttcacatggtttgaatcacccaggccgcacagagatgagatgccttta	1447
Db	350	GTACTCATGGGTTTCACATGGTTTGAATCACAGGGCCGCCATCAGGAGATGCTTTA	291
Qy	1448	tccagatgaagtctcgacagacatttatgctgcacagaagtgtcataaaaaaaca	1507
Db	290	TCCAGATGAAGTCTCGGACAGACATTTATGCGTGCACAGAAGTGTTCAT-AAAAAACA	232
Qy	1508	tgaagcacagatatgttaagtcttctcagttcagctgagagatgaactttgtttaa	1567
Db	231	TGAAGCACAGATATGTTGAAGTCTTTCAGTGTTCAGTGAGGAGATGAACTTTGTTAA	172
Qy	1568	tggggggcacttaaatcgaaatggcttattccccccg-----ccatgcctgt	1615









COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/08/764,100  
APPLICATION NUMBER: US/08/764,100  
FILING DATE: 06-DEC-1996  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/214,064  
FILING DATE:  
APPLICATION NUMBER: US 08/032,235  
FILING DATE: 17-MAR-1993  
APPLICATION NUMBER: GB 9206016.9  
FILING DATE: 19-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5773700ris, Allen E.  
REGISTRATION NUMBER: 34,490  
REFERENCE/DOCKET NUMBER: 137-1061  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 354-3592  
TELEFAX: (415) 857-1125  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2993 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-764-100-10

Query Match 1.3%; Score 47.8; DB 1; Length 2993;  
Best Local Similarity 47.8%; Pred. No. 0.0073;  
Matches 139; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

Qy 2570 atgatatctttagtataaaacaaataaattttaaaccacctgaagataaattagaaga 2629  
Db 1494 ATCATATGAATAAATAAATAACATAAATAAATAAATAAATAAATAAATAAATAA 1435  
Qy 2630 aattgtgacccctccacaaacatacaaaagttaaagtttgatcttttcttcagcagg 2689  
Db 1434 TAATAAAACCAAAAAGGATGGCCCTCGGGCACAAATTTGGTTGCTTTAATAATGCTTT 1375  
Qy 2690 tatcagttgtaataatgaattagggccaaatgcacaaatgcacaaatgcacagctaca 2749  
Db 1374 AAAATGAATGATTAAGTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1315  
Qy 2750 tgtagttagtaattcttagttgactgaattgaataattgtggtcttcattattatt 2809  
Db 1314 AATTGTATTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTT 1255  
Qy 2810 ttattgttacttttttcttattgatgggttgacatttaagagagaaat 2860  
Db 1254 TTTTATTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTT 1204

RESULT 8  
US-08-764-100-1  
Sequence 1, Application US/08764100  
Patent No. 5773700  
GENERAL INFORMATION:  
APPLICANT: van Grinsven J., Martinus Q.  
APPLICANT: De Haan, Petrus T.  
APPLICANT: Gielen L., Johannes J.  
APPLICANT: Peters, Dirk  
APPLICANT: Goldbach, Robert W.  
TITLE OF INVENTION: Improvements in or Relating to Organic  
TITLE OF INVENTION: Compounds  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sandoz Agro, Inc.  
STREET: 975 California Avenue  
CITY: Palo Alto  
STATE: CA

COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/764,100  
FILING DATE: 06-DEC-1996  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/214,064  
FILING DATE:  
APPLICATION NUMBER: US 08/032,235  
FILING DATE: 17-MAR-1993  
APPLICATION NUMBER: GB 9206016.9  
FILING DATE: 19-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5773700ris, Allen E.  
REGISTRATION NUMBER: 34,490  
REFERENCE/DOCKET NUMBER: 137-1061  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 354-3592  
TELEFAX: (415) 857-1125  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3001 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-764-100-1

Query Match 1.3%; Score 47.8; DB 1; Length 3001;  
Best Local Similarity 47.8%; Pred. No. 0.0073;  
Matches 139; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

Qy 2570 atgatatctttagtataaaacaaataaattttaaaccacctgaagataaattagaaga 2629  
Db 1508 ATCATATGAATAAATAAATAACATAAATAAATAAATAAATAAATAAATAAATAA 1567  
Qy 2630 aattgtgacccctccacaaacatacaaaagttaaagtttgatcttttcttcagcagg 2689  
Db 1568 TAATAAAACCAAAAAGGATGGCCCTCGGGCACAAATTTGGTTGCTTTAATAATGCTTT 1627  
Qy 2690 tatcagttgtaataatgaattagggccaaatgcacaaatgcacaaatgcacagctaca 2749  
Db 1628 AAAATGAATGATTAAGTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1687  
Qy 2750 tgtagttagtaattcttagttgactgaattgaataattgtggtcttcattattatt 2809  
Db 1688 AATTGTATTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTT 1747  
Qy 2810 ttattgttacttttttcttattgatgggttgacatttaagagagaaat 2860  
Db 1748 TTTTATTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTT 1798

RESULT 9  
US-08-998-416-288/c  
Sequence 288, Application US/08998416  
Patent No. 6239264  
GENERAL INFORMATION:  
APPLICANT: Philippsen, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jurgen  
APPLICANT: Knechtie, Philipp  
APPLICANT: Rebeschung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPPII  
TITLE OF INVENTION: AND USES THEREOF





```
; NAME: No. 5773700ris, Allen E.
; REGISTRATION NUMBER: 34,490
; REFERENCE/DOCKET NUMBER: 137-1061
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 354-3592
; TELEFAX: (415) 857-1125
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3000 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-764-100-9

Query Match          1.2%; Score 43.8; DB 1; Length 3000;
Best Local Similarity 49.1%; Pred. No. 0.083;
Matches 143; Conservative 0; Mismatches 147; Indels 1; Gaps 1;

QY 2570 atgatattcttgattgataaaacacaaatttaaaacacacctgaagataaattagaaga 2629
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1493 ATCATATGAATAAATATAAATAACATAAAATATAAATAAATACAAAAAATCATAAAA 1434

QY 2630 aatttgaccctccacaaaacatacaaaagtttaaaagtttgatctttttctcagcagg 2689
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1433 TAAATATAAAACCAAAAAGGATGGCTTCGGGCACAA-TTTGGTTGCTTTTAAATAATGCTT 1375

QY 2690 tatcagttgtaaatgaattaggggcacaaatgcacaaacgacaaatgaagcagctaca 2749
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1374 TAAAGAATGATATTAGTAAATTAATACTTTAAATCCCAATCTACACAAATGGCCAAA 1315

QY 2750 tgaagttagtaattctctagttgaaactgtaattgaaattggtggttcattgtattatt 2809
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1314 AATTTGTAATGCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1255

QY 2810 ttatatgtactttttcattattgatggtttgggactttaataagagagaat 2860
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1254 TTTTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1204

RESULT 13
US-07-867-106-2
; Sequence 2, Application US/07867106
; Patent No. 5389526
;
GENERAL INFORMATION:
; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
;
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/867,106
; FILING DATE: 19920625
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PJ 7187
; APPLICATION NUMBER: PCT/AU90/00530
; FILING DATE: 02-NOV-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Feeney, Joanne Longo
; REGISTRATION NUMBER: 35,134
```

```
; REFERENCE/DOCKET NUMBER: RICE-0002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5852 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
;
US-07-867-106-2

Query Match          1.2%; Score 43.8; DB 1; Length 5852;
Best Local Similarity 44.8%; Pred. No. 0.11;
Matches 168; Conservative 0; Mismatches 207; Indels 0; Gaps 0;

QY 2583 attaaaaacaaatttaaaacacacctgaagataaattagaagaaattgtgcacct 2642
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5387 ATTTATAAAATTTGAAGTTCATCAAGATATATAGATAAATTAATTAATTAATTTGAATTTT 5446

QY 2643 ccacaaaacatacaaaagtttaaaagtttggatctttttctcagcaggtatcagttgtaaa 2702
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5447 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 5506

QY 2703 taatgaattaggggcacaaatgcacaaacgacaaatgaagcagctacatgtagttagtaat 2762
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5507 TCTGCTCAATGATTTTAAATAAATAAATCGATACATAAATTTTAAAAAACCCTTTACATT 5566

QY 2763 ttctagtttgaactgtaattgaaataattgtggttcattgtattatttttatttattt 2822
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5567 TTTTATTTTAAATCCAAATTTATACATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 5626

QY 2823 ttttcattatgattggtttggactttaataagagaaattccaatgttttttaataatccag 2882
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5627 TTAATTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 5686

QY 2883 aagtgagacaattgacagtgatctctagaaaaacatacactaacctgaacagaagtga 2942
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5687 ATAAATAAAAAATACAAATATAAGTAAAAAACAACAAACAAATACATATATATAAAAAA 5746

QY 2943 tgcctatatattat 2957
   ||||| ||||| |||||
Db 5747 TACAAATAACAAATA 5761

RESULT 14
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;
GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```



```
Db 178 ATAAATATTAAATGAATGAATTAAGTAAATATATAATAATAATAAGTATTAAT 119
QY 2893 attgacacagtgtattctagaaacaatacacactaacctgaacagaagtgaatgcttatata 2952
Db 118 AATCAAAATAATTAAATTTATTAAATAATGATAATAATAGTTTAAATAAATTAATACTTAAA 59
QY 2953 tattatgatagccttaaacctttttcctcctaatgccttaactgtcaataaattataac 3010
Db 58 TATTATAATAAAAAAGTTTATATTAATCTTTATAAATTAAATTATATATATTAGATC 1
```

Search completed: August 15, 2002, 14:12:13  
Job time: 29244 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2002, 06:20:49 ; Search time 2423.18 seconds  
(without alignments)  
2540.108 Million cell updates/sec

Title: US-09-697-206A-1  
Perfect score: 3585  
Sequence: 1 cttttgggacactgctgg.....gcctgttaatctcaaaaaa 3585

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues  
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_032802.\*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3585	100.0	3598	22 AAF85701	Human cancer relat
2	2521.4	70.3	2882	21 AAA12412	cDNA encoding a hu
3	2360	65.8	2731	22 AAF98702	Human ovarian-canc
4	2316.8	64.6	2468	22 AAH33714	Human colon cancer
5	1986.2	55.4	2261	21 AAC77930	Human cancer assoc
6	1324.8	37.0	1624	22 AAH14740	Human cDNA sequenc
7	587.4	16.4	1585	21 AAC59113	Human secreted pro
8	555.6	15.5	740	20 AAZ16154	Human gene express
9	524	14.6	745	20 AAZ16155	Human gene express

c	10	505.4	14.1	565	22	AAH12166	Human cDNA clone (
	11	498.4	13.9	503	22	AAF68133	Human lung tumour
	12	498.4	13.9	503	22	AAF68134	Human lung tumour
	13	441.4	12.3	700	21	AAA01641	Human colon cancer
	14	392.4	10.9	745	22	AAH07171	Human cDNA clone (
	15	351	9.8	353	22	AAI29269	Colon tumour relat
	16	343.8	9.6	396	22	AAF94890	Human ovarian canc
	17	322.8	9.0	329	22	AAI29175	Colon tumour relat
c	18	305	8.5	351	21	AAZ80296	Human colon cancer
	19	272.4	7.6	411	22	AAZ80296	cDNA encoding nove
	20	272.4	7.6	411	22	AAK87738	Human digestive sy
c	21	271	7.6	271	22	AAK87738	Human breast cance
c	22	271	7.6	271	22	AAK87738	Human breast cance
	23	231	6.4	300	20	AAZ14741	Human gene express
	24	227	6.3	2115	23	ABL12241	Drosophila melanog
c	25	223	6.2	294	22	AAF98661	Human ovarian can
	26	166	4.6	241	22	AAZ46997	Human breast cance
	27	166	4.6	241	22	AAZ46997	Human breast cance
	28	156.6	4.4	260	21	AAA00145	Human colon cancer
	29	153	4.3	153	22	AAZ39650	Genomic sequence #
	30	153	4.3	153	22	AAZ39650	Human digestive sy
	31	151.4	4.2	157	16	AAZ4184	Human gene signatu
c	32	129	3.6	155	21	AAA45873	Human secreted exp
	33	127.4	3.6	9102	23	ABL12240	Drosophila melanog
	34	120.2	3.4	317	20	AAZ40462	Human secreted pro
	35	118.2	3.3	339	23	AAZ58815	cDNA #1491 encodin
	36	116.8	3.3	427	21	AAZ76005	Human OREX ORF1560
	37	79.8	2.2	242	22	AAZ13980	Human breast cance
	38	79.4	2.2	257	22	AAZ14219	Human breast cance
	39	66.4	1.9	2670	22	AAH14219	Human cDNA sequenc
	40	61.4	1.7	826	22	AAI97495	Human neuroblastom
	41	61.2	1.7	1036	21	AAZ00757	Human secreted pro
	42	60.2	1.7	1034	21	AAZ10857	Human secreted pro
	43	60.2	1.7	2201	22	AAZ83093	HMRP H1 ovarian tu
	44	60	1.7	815	22	AAI96163	Human neuroblastom
	45	58.6	1.6	2234	23	AAZ66261	DNA encoding novel

## ALIGNMENTS

RESULT	1
AAF85701	
ID	AAF85701 standard; cDNA; 3598 BP.
XX	
AC	AAF85701;
XX	
DT	10-DEC-2001 (first entry)
XX	
DE	Human cancer related protein 20P2H8 coding sequence.
XX	
KW	Human; cancer related protein 20P2H8; vaccine; chromosome 15q32-23;
KW	prostate cancer; bladder cancer; colon cancer; pancreatic cancer; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	1..2148
FT	/*tag= a
FT	/product= "20P2H8 alternative version"
FT	451..2004
FT	/*tag= b
FT	/product= "20P2H8"
XX	
PN	WO200131012-A1.
XX	
PD	03-MAY-2001.
XX	
PF	26-OCT-2000; 2000WO-US29477.
XX	
PR	28-OCT-1999; 99US-0162364.
XX	
PA	(UROC-) UROGENESYS INC.



QY 1801 cttgctactctccacagctgctgaattcttagcgtgtccctccacagcagctggcagcgtg 1860  
Db 1801 cttggctactctccacagctgctgaattcttagcgtgtccctccacagcagctggcagcgtg 1860  
QY 1861 gtcaagaatgcaggcctggcctacaaactgagtgagtgaaagaaatcttaactttctcaa 1920  
Db 1861 gtcaagaatgcaggcctggcctacaaactgagtgagtgaaagaaatcttaactttctcaa 1920  
QY 1921 ggttaccagtgatgcacagcagagtgagtgatatacacacaaatgaccagggcagagactcta 1980  
Db 1921 ggttaccagtgatgcacagcagagtgagtgatatacacacaaatgaccagggcagagactcta 1980  
QY 1981 cccaagaatggtttgtatttaagggcccgagcagtgatgaacatctccagaaagaagt 2040  
Db 1981 cccaagaatggtttgtatttaagggcccgagcagtgatgaacatctccagaaagaagt 2040  
QY 2041 gtttgaagatgatgttgatcttgaaacctccagacacagaaacactcttagcaaatc 2100  
Db 2041 gtttgaagatgatgttgatcttgaaacctccagacacagaaacactcttagcaaatc 2100  
QY 2101 aggggaagtgtgtcacactcagcgtcagtgattttcagcaaaactgttggaacaaacgg 2160  
Db 2101 aggggaagtgtgtcacactcagcgtcagtgattttcagcaaaactgttggaacaaacgg 2160  
QY 2161 gctgtgcttatcttttgggtgagtgaaataatttgagctagtgaaagccaaatcgtaac 2220  
Db 2161 gctgtgcttatcttttgggtgagtgaaataatttgagctagtgaaagccaaatcgtaac 2220  
QY 2221 ttacagcaagcagatgcagcacaactggcgtcttcttgctgattgcaaaataggcatttaaaa 2280  
Db 2221 ttacagcaagcagatgcagcacaactggcgtcttcttgctgattgcaaaataggcatttaaaa 2280  
QY 2281 tgtgaatttgaatcagatgtctccacttacttccagtttcaagtgagcgcatacagtggttcc 2340  
Db 2281 tgtgaatttgaatcagatgtctccacttacttccagtttcaagtgagcgcatacagtggttcc 2340  
QY 2341 ctaagtttaagtcttggtataaaactccacagtgctacacatccaccatgaactct 2400  
Db 2341 ctaagttttaagtcttggtataaaactccacagtgctacacatccaccatgaactct 2400  
QY 2401 gtttaggaagcttcaattttgtatattccgcgtcttctctcttcttcaattccctgtctctg 2460  
Db 2401 gtttaggaagcttcaattttgtatattccgcgtcttctctcttcttcaattccctgtctctg 2460  
QY 2461 cataatcatgctctcttctgtaagtaattcaagcacaagaatcttgaataataaaatcaca 2520  
Db 2461 cataatcatgctctcttctgtaagtaattcaagcacaagaatcttgaataataaaatcaca 2520  
QY 2521 atcttaggaaagaataaaattgttattttcccgagtccttggccatgatgatattctta 2580  
Db 2521 atcttaggaaagaataaaattgttattttcccgagtccttggccatgatgatattctta 2580  
QY 2581 tgattaaaaaacaataattttaaacaactgaagataaaattagaagaataattgtgcacc 2640  
Db 2581 tgattaaaaaacaataattttaaacaactgaagataaaattagaagaataattgtgcacc 2640  
QY 2641 cttcccaaaaacatacaaaagttaaagtttgatcttttctcagcaggtatcagttgta 2700  
Db 2641 cttcccaaaaacatacaaaagttaaagtttgatcttttctcagcaggtatcagttgta 2700  
QY 2701 aataatgaattaggggcccataatgcacaaacgaaataatgaagcagctacatagtagtagta 2760  
Db 2701 aataatgaattaggggcccataatgcacaaacgaaataatgaagcagctacatagtagtagta 2760  
QY 2761 atttctagttgaactgtaattgaatattgtggttccatgtattatttttatattgtac 2820  
Db 2761 atttctagttgaactgtaattgaatattgtggttccatgtattatttttatattgtac 2820  
QY 2821 ttttttctattatgtatggttggacttttaataagagaataatccatagtttttaatatccc 2880  
Db 2821 ttttttctattatgtatggttggacttttaataagagaataatccatagtttttaatatccc 2880  
QY 2881 agaagtgcacaatttgaacagtgatttctagaaaaacaatacactaaactgaacagaagt 2940

Db 2881 agaagtgcacaatttgaacagtgatttctagaaaaacaatacactaaactgaacagaagt 2940  
QY 2941 aatgcttatatatattatgatagccttaaaccttttctcttaactgaacttaactgcaaa 3000  
Db 2941 aatgcttatatatattatgatagccttaaaccttttctcttaactgaacttaactgcaaa 3000  
QY 3001 taattataaacttttaagcattagactatagtcagcagctagactgagaggttaaacac 3060  
Db 3001 taattataaacttttaagcattagactatagtcagcagctagactgagaggttaaacac 3060  
QY 3061 tgaatcaattagaacaggtactgactgctgagtggttttaacacactatgttttagctgtgtt 3120  
Db 3061 tgaatcaattagaacaggtactgactgctgagtggttttaacacactatgttttagctgtgtt 3120  
QY 3121 atgctataaaagtgaataattatagacactagtagtactgctgctcctcactgaactccaaag 3180  
Db 3121 atgctataaaagtgaataattatagacactagtagtactgctgctcctcactgaactccaaag 3180  
QY 3181 aaaaacaggtattcattaaagtgcattgaatgtgataatttcttaagttaactatattgtc 3240  
Db 3181 aaaaacaggtattcattaaagtgcattgaatgtgataatttcttaagttaactatattgtc 3240  
QY 3241 ctttcttgtaatgcaatccgtgcagatttatgagctgctatttttttcttctgtgcac 3300  
Db 3241 ctttcttgtaatgcaatccgtgcagatttatgagctgctatttttttcttctgtgcac 3300  
QY 3301 tactttaacaccttaaaagggaagaacaaacatttctctcagctgactgagcgaatggcc 3360  
Db 3301 tactttaacaccttaaaagggaagaacaaacatttctctcagctgactgagcgaatggcc 3360  
QY 3361 ctttaactgcgaataggaagaaaaaaaagggttggtaaaattggtaaacctggca 3420  
Db 3361 ctttaactgcgaataggaagaaaaaaaagggttggtaaaattggtaaacctggca 3420  
QY 3421 ctttaagatcgaaaaaatttctgtatacttgcagtccttaagatgcccaagctgccccaa 3480  
Db 3421 ctttaagatcgaaaaaatttctgtatacttgcagtccttaagatgcccaagctgccccaa 3480  
QY 3481 agctctgaagaactttaagatagaggcagtaagtcttaactaactaactagtttttga 3540  
Db 3481 agctctgaagaactttaagatagaggcagtaagtcttaactaactaactagtttttga 3540  
QY 3541 gaggttcaacatttgataataaaactgctgtttaaactcacaataa 3585  
Db 3541 gaggttcaacatttgataataaaactgctgtttaaactcacaataa 3585

## RESULT 2

AAAL2412

ID AAAL2412 standard; cDNA; 2882 BP.

XX AAAL2412;

XX 25-JUL-2000 (first entry)

XX cDNA encoding a human RNA-associated protein.

DE Human; RNA-associated protein; cell proliferation; cancer; inflammation;  
immune response; reproductive disorder; actinic keratosis;  
KW atherosclerosis; arteriosclerosis; bursitis; cirrhosis; hepatitis;  
KW mixed connective tissue disease; myelofibrosis; primary thrombocythemia;  
KW paroxysmal nocturnal hemoglobinuria; polycythemia vera; psoriasis;  
KW trauma; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FH Key 293..2338

FT CDS /\*tag= a

FT /product= "RNA-associated protein"

XX PN WO200015799-A2.

```
XX 23-MAR-2000.
XX 17-SEP-1999; 99WO-US21688.
XX 17-SEP-1998; 98US-0156039.
XX 22-SEP-1998; 98US-0158720.
XX 04-NOV-1998; 98US-0186815.
XX 08-APR-1999; 99US-0128660.
XX (INCY-) INCYTE PHARM INC.
XX Tang YT, Corley NC, Guegler KJ, Gorgone GA, Patterson C;
XX Hillman JL, Baughn MR, Lal P, Azimzai Y, Yue H, Yang J;
XX WPI; 2000-271437/23.
XX P-PSDB; RAY8440.
XX
XX New polypeptides and polynucleotides, useful for preventing and
XX treating a disorder associated with increased or decreased expression
XX of RNA associated proteins -
XX
XX Claim 9; Page 121-122; 131pp; English.
XX
XX The present sequence encodes a human RNA-associated protein. The
XX expression of RNA-associated proteins is closely associated with
XX reproductive tissues, nervous tissues, cell proliferation including
XX cancer, inflammation and immune responses, and so they may be used
XX for diagnosis, treatment or prevention of cell proliferative,
XX immune/inflammatory disorders, and reproductive disorders. Diseases
XX and disorders which may be treated include actinic keratosis,
XX atherosclerosis, arteriosclerosis, bursitis, cirrhosis, hepatitis,
XX mixed connective tissue disease, myelofibrosis, paroxysmal nocturnal
XX hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia
XX and cancers, and trauma.
XX
XX Sequence 2882 BP; 755 A; 691 C; 671 G; 765 T; 0 other;
XX
Query Match 70.3%; Score 2521.4; DB 21; Length 2882;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2544; Conservative 0; Mismatches 1; Indels 12; Gaps 1;
Qy 1 ctttttggatcaactgtggtggccaccgggcccagctagctcgatgagaagattg 60
Db 323 ctttttggatcaactgtggtggccaccgggcccagctagctcgatgagaagattg 382
Qy 61 atcctgctgttctgaaagtcgtggtgctgcccacaaagaggtggacagttgcacgaa 120
Db 383 atcctgctgttctgaaagtcgtggtgctgcccacaaagaggtggacagttgcacgaa 442
Qy 121 gtgctagttagaccggtacgttggaactgacggagactgcaagaagaactaaata 180
Db 443 gtgctagttagaccggtacgttggaactgacggagactgcaagaagaactaaata 502
Qy 181 gacgtcaaaagcctgtctctggcgtcagctggaccagccctccgacagtttaaccag 240
Db 503 gacgtcaaaagcctgtctctggcgtcagctggaccagccctccgacagtttaaccag 562
Qy 241 tcagtgaacatgaactgaatattggagtaggaggtcctctctctgtactgatggg 300
Db 563 tcagtgaacatgaactgaatattggagtaggaggtcctctctctgtactgatggg 622
Qy 301 cagcttcagtgcaggcaaaactcctgcatcctgaggtctccaaagaagaatgtactattcct 360
Db 623 cagcttcagtgcaggcaaaactcctgcatcctgaggtctccaaagaagaatgtactattcct 682
Qy 361 gaatgcttctattccttttttggatcttcgaaaagaataatcgaagaattgcccctgttca 420
Db 683 gaatgcttctattccttttttggatcttcgaaaagaataatcgaagaattgcccctgttca 742
Qy 421 cctgatattgacaactggagctggccacaatgacagattatttaatttttgagaagagt 480
Db 1823 aaaaacatgaaggacagatattgtgaaagtccttcagtgctcagctggaggagatgaacttt 1882
```

```
Db 743 cctgatattgacaaaactggacgtgtgcccacaatgacagagattttaatttttgagaagagt 802
Qy 481 agttcagttctcgtatgtaggacctctcaagtgaagatatgggaatatatttttagca 540
Db 803 agttcagttctcgtatgtaggacctctcaagtgaagatatgggaatatatttttagca 862
Qy 541 atgatttcagagccttataatcacaggtttttcagatccagagagagtgaaattacaagt 600
Db 863 atgatttcagagccttataatcacaggtttttcagatccagagagagtgaaattacaagt 922
Qy 601 gaaagtgaacttcagacaagatggaacttattgatataacacccgtagtcagggcacga 660
Db 923 gaaagtgaacttcagacaagatggaacttattgatataacacccgtagtcagggcacga 982
Qy 661 ggtttaccatggcagctcttcagatcaagatatttcaagattcttcaaggagactcaatt 720
Db 983 ggtttaccatggcagctcttcagatcaagatatttcaagattcttcaaggagactcaatt 1042
Qy 721 gccaaaggaggtgcagcacctttgtcctaagtctcaggttcgaaagcggagagagctctg 780
Db 1043 gccaaaggaggtgcagcacctttgtcctaagtctcaggttcgaaagcggagagctctg 1102
Qy 781 gttaggttttaagtggagcaccgagacccagactagcactacagagggcacaacatcacatg 840
Db 1103 gttaggttttaagtggagcaccgagacccagactagcactacagagggcacaacatcacatg 1162
Qy 841 gggaccgcgttatattgaggttttacaagcacaaggttgaagatttcccttaaatgtcgtg 900
Db 1163 gggaccgcgttatattgaggttttacaagcacaaggttgaagatttcccttaaatgtcgtg 1222
Qy 901 ggtactccaatgaggttagccagtttctctccaaagaaatacaagtcattgttcgcatg 960
Db 1223 ggtactccaatgaggttagccagtttctctccaaagaaatacaagtcattgttcgcatg 1282
Qy 961 cgggggctcctcttcacgcgcacagctgaagaagtgtggcctcttttggacacagcttgc 1020
Db 1283 cgggggctcctcttcacgcgcacagctgaagaagtgtggcctcttttggacacagcttgc 1342
Qy 1021 cctattactgggggaaaggagcgcacccctctttgtcacctaccacagatggtagccaca 1080
Db 1343 cctattactgggggaaaggagcgcacccctctttgtcacctaccacagatggtagccaca 1402
Qy 1081 ggggacgcttttgcctctcttgcctgtgaggaatatgcacagaatgctgtgaggaagcat 1140
Db 1403 ggggacgcttttgcctctcttgcctgtgaggaatatgcacagaatgctgtgaggaagcat 1462
Qy 1141 aaagactgtgtggtgaaagatacatgtaactcttcagagcacagcagctgaagttcag 1200
Db 1463 aaagactgtgtggtgaaagatacatgtaactcttcagagcacagcagcagctgaagttcag 1522
Qy 1201 caggtgctgaatcgtattcctcctggccctctcattccacttccaaacccctccattatt 1260
Db 1523 caggtgctgaatcgtattcctcctggccctctcattccacttccaaacccctccattatt 1582
Qy 1261 ccagttactacctaagaatttgtgcccctacaaaatgttagagactgtatacgccttcga 1320
Db 1583 ccagttactacctaagaatttgtgcccctacaaaatgttagagactgtatacgccttcga 1642
Qy 1321 ggtctccctcattgagccacaatgagacatcctgattcctcctggggaggttcgcacaa 1380
Db 1643 ggtctccctcattgagccacaatgagacatcctgattcctcctggggaggttcgcacaa 1702
Qy 1381 gatattcgtactcatgggtttcacatggtttgaaatcacccagggcccccacatcaggagat 1440
Db 1703 gatattcgtactcatgggtttcacatggtttgaaatcacccagggcccccacatcaggagat 1762
Qy 1441 gcctttatccagatgaagttcgtgcgacagagcatttatgctgcacagaagtgctataaa 1500
Db 1763 gcctttatccagatgaagttcgtgcgacagagcatttatgctgcacagaagtgctataaa 1822
Qy 1501 aaaaacatgaagcacagatattgtgaagctcttcagtgctcagctggaggagatgaacttt 1560
Db 1823 aaaaacatgaagcacagatattgtgaagctcttcagtgctcagctggaggagatgaacttt 1882
```



Db 256 ||||| ctggtgtaactccaatgagtgagccagtttctctccaaggaaaatcaagtcatgtcc 315  
 QY 956 gcatcgggggctccctttcacggccacagctgaagaagtgggtggtctctttggacagc 1015  
 Db 316 gcatcgggggctccctttcacggccacagctgaagaagcgggtggtctctttggacagc 375  
 QY 1016 attgcccattactgggggaaaggaaaggaatcctctttgtcacctaccagatggtaggc 1075  
 Db 376 attgcccattactgggggaaaggaaaggaatcctctttgtcacctaccagatggtaggc 435  
 QY 1076 caacagggagcgttttgcctctttgtcctgtgaggaataatgacagaaatgcgttagga 1135  
 Db 436 caacagggagcgttttgcctctttgtcctgtgaggaataatgacagaaatgcgttagga 495  
 QY 1136 agcataaagacttgggttaaaagatacattgaactcttcaggagacagcagctgaag 1195  
 Db 496 agcataaagacttgggttaaaagatacattgaactcttcaggagacagcagctgaag 555  
 QY 1196 ttccagcaggtgctgaatcgattctctcgccctctctcattccacttccaaacctccca 1255  
 Db 556 ttccagcaggtgctgaatcgattctctcgccctctctcattccacttccaaacctccca 615  
 QY 1256 ttattccagtaactacagcaatttgcctccctcaaaatgtagagactgtatacgc 1315  
 Db 616 ttattccagtaactacagcaatttgcctccctcaaaatgtagagactgtatacgc 675  
 QY 1316 ttccaggtctccctatgcagccacaaatgaggacatcctcggttccctggggaggtcg 1375  
 Db 676 ttccaggtctccctatgcagccacaaatgaggacatcctcggttccctggggaggtcg 735  
 QY 1376 ccacagatacttgtaactcgtgggttcacatggttttgaaaccaggggcccccacag 1435  
 Db 736 ccacagatacttgtaactcgtgggttcacatggttttgaaaccaggggcccccacag 795  
 QY 1436 gagatccctttaccagatgaagcttcgggacagacatttatggctgacagaagtgtc 1495  
 Db 796 gagatccctttaccagatgaagcttcgggacagacatttatggctgacagaagtgtc 855  
 QY 1496 ataaaaaaacatgaaggacagatgttgaaatccttccagtttcagcttcagctgagagatga 1555  
 Db 856 ataaaaaaacatgaaggacagatgttgaaatccttccagtttcagcttcagctgagagatga 915  
 QY 1556 actttgttaattgggggacgtttaaatacgaatggcttatccccacgcg----- 1605  
 Db 916 actttgttaattgggggacgtttaaatacgaatggcttatccccacgcgccaatgaagt 975  
 QY 1606 --ccatgctgtctcctcctcctacacatttccagctcctgctgagttattcctacag 1663  
 Db 976 taccatgctgtctcctcctcctcctacacatttccagctcctgctgagttattcctacag 1035  
 QY 1664 aagctgccatttaccagccctctgtgatgttgaaatccagagcactgcagccctccacag 1723  
 Db 1036 aagctgccatttaccagccctctgtgatgttgaaatccagagcactgcagccctccacag 1095  
 QY 1724 cgtactaccagcagcactcagctcttcattgaactacacagcgtactatccccagcccc 1783  
 Db 1096 cgtactaccagcagcactcagctcttcattgaattacacagcgtactatccc----- 1148  
 QY 1784 caggttcgcctaatagtcttggtcacttccctacagctcctaacttagcgggtgtccctc 1843  
 Db 1149 ----- 1148  
 QY 1844 cacagctggcacggtggtcagaatgcagggcctgacctacaatactgggttaaggaaa 1903  
 Db 1149 ----- 1148  
 QY 1904 ttcttaacttctccaaggttaccagtatgcaaccgaggtgagcttatcacacaaatg 1963  
 Db 1149 ----- 1148  
 QY 1964 accagccaggaactctacccaaagaatgggtttgtatttaaggggcccccagcagttagaac 2023

Db 1149 ----- 1148  
 QY 2024 atcctcagaaaagaagtgtttgaaagatgtatggtgatcttgaaacctccagacacaaga 2083  
 Db 1149 ----- agtgtttgaaagatgtatggtgatcttgaaacctccagacacaaga 1194  
 QY 2084 aaacttctagcaaatccaggggaagtttgtcacactcaggctgcagatattttcagcaaa 2143  
 Db 1195 aaacttctagcaaatccaggggaagtttgtcacactcaggctgcagatattttcagcaaa 1254  
 QY 2144 ctgatttgacaacacggcctgtgccttatcttttgttgagtgaaaaaatttgagctag 2203  
 Db 1255 ctgatttgacaacacggcctgtgccttatcttttgttgagtgaaaaaatttgagccag 1314  
 QY 2204 tgaagccaaatcgtaaacttacagcaagcagcatgcagcatcacctggctctttgctgattg 2263  
 Db 1315 tgaagccaaatcgtaaacttacagcaagcagcatgcagcatcacctggctctttgctgattg 1374  
 QY 2264 caaataggcatttaaaatgtgaatttggaaatcagatgtctccattacttccagttaaagt 2323  
 Db 1375 caaataggcatttaaaatgtgaatttggaaatcagatgtctccattacttccagttaaagt 1434  
 QY 2324 ggcatcataggtgtttcctaaagttttaagctcttgataaaaaactccaccagttctacca 2383  
 Db 1435 ggcatcataggtgtttcctaaagttttaagctcttgataaaaaactccaccagttctacca 1494  
 QY 2384 tctccaccatgaactctgttaaggaaagcttcaattttgttatattcccgctctttctctt 2443  
 Db 1495 tctccaccatgaactctgttaaggaaagcttcaattttgttatattcccgctctttctctt 1554  
 QY 2444 cattccctgtctcttgcataatcatcgtccttcttgcataatcgaagtaattcgaagcataagatt 2503  
 Db 1555 cattccctgtctcttgcataatcatcgtccttcttgcataatcgaagtaattcgaagcataagatt 1614  
 QY 2504 ggaataataaaatcacaaatcttagggagaaagaataaaattgttatatttccagctctcttg 2563  
 Db 1615 ggaataataaaatcacaaatcttagggagaaagaataaaattgttatatttccagctctcttg 1674  
 QY 2564 gccatgatgatacttatgattaaaaacaaatlaaaattttaaaccacctgaagataaatt 2623  
 Db 1675 gccatgatgatacttatgattaaaaacaaatlaaaattttaaaccacctgaagataaatt 1734  
 QY 2624 agaagaaattgtcacccctccacaaacatacacaagtttaaaagtttggaattctttctc 2683  
 Db 1735 agaagaaattgtcacccctccacaaacatacacaagtttaaaagtttggaattctttctc 1794  
 QY 2684 acagatatacagttgataataatgaattaggggccaaaatgcaaaacgaaaaatgaagca 2743  
 Db 1795 agcaggtatcagttgataataatgaattaggggccaaaatgcaaaacgaaaaatgaagca 1854  
 QY 2744 gctacatgtagttagtaaatcttagtttgaaactgtaattgaattgtggtctcatagt 2803  
 Db 1855 gctacatgtagttagtaaatcttagtttgaaactgtaattgaattgtggtctcatagt 1914  
 QY 2804 attattttatattgtaacttttttccattattgattggtttggaactttaagaagaaattcc 2863  
 Db 1915 attattttatattgtaacttttttccattattgattggtttggaactttaagaagaaattcc 1974  
 QY 2864 atagtttttaataatccagaaagagacaatttgaacaggtattcttagaaaaacaataca 2923  
 Db 1975 atagtttttaataatccagaaagagacaatttgaacaggtattcttagaaaaacaataca 2034  
 QY 2924 ctaactgaacagaaagtgaattgcttatataattatgatagccttaaaccttttccctcta 2983  
 Db 2035 ctaactgaacagaaagtgaattgcttatataattatgatagccttaaaccttttccctcta 2094  
 QY 2984 atgccttaactgtcacaataattataaccttttaagcataggactatagtcagcatgcta 3043  
 Db 2095 atgccttaactgtcacaataattataaccttttaagcataggactatagtcagcatgcta 2154  
 QY 3044 gactgagaggttaaacactgatcaattagacaggtactgatgctgctcaggttttaacac 3103  
 Db 2155 gactgagaggttaaacactgatcaattagacaggtactgatgctgctcaggttttaacac 2214

QY 3104 tatgtttagctgtttatgctataaaagtgcataattagacactagctagctactgctgc 3163  
|||||  
Db 2215 tatgtttagctgtttatgctataaaagtgcataattagacactagctagctactgctgc 2274  
QY 3164 ctcagttaactccaaagaacagagatttcattaaagtgcattgaatggtgataattctct 3223  
Db 2275 ctcagttaactccaaagaacagagatttcattaaagtgcattgaatggtgataattctct 2334  
QY 3224 agttactcatgttccttcttgatgcaatgcgatgcgctgcagatttatgagctgctat 3283  
|||||  
Db 2335 agttactcatgttccttcttgatgcaatgcgatgcgctgcagatttatgagctgctat 2394  
QY 3284 ttttattttctgtcattactttaacaccttaagggagagacaaacatttctcttcca 3343  
|||||  
Db 2395 ttttattttctgtcattactttaacaccttaagggagagacaaacatttctcttcca 2454  
QY 3344 gctgactgcaatgccccttttaactgcaataggaagaaacaaaggtttgtgaaa 3403  
Db 2455 gctgactgcaatgccccttttaactgcaataggaagaaacaaaggtttgtgaaa 2514  
QY 3404 attggtgataactggcacctaaagatcgaaagaaattctgtatctatgaccttaaga 3463  
|||||  
Db 2515 attggtgataactggcacctaaagatcgaaagaaattctgtatctatgaccttaaga 2574  
QY 3464 tgcccacaaagctgccaaagctctgaaagactttaagatagggcagtaagtcttactacaat 3523  
|||||  
Db 2575 tgcccacaaagctgccaaagctctgaaagactttaagatagggcagtaagtcttactacaat 2634  
QY 3524 actactgagttttgtagagtttaacatttgataataaaacttgcctgtttaatctcaaaa 3583  
Db 2635 actactgagttttgtagagtttaacatttgataataaaacttgcctgtttaatctcaaaa 2694  
QY 3584 aa 3585  
Db 2695 aa 2696  
RESULT 4  
AAH33714  
ID AAH33714 standard; cDNA; 2468 BP.  
XX  
AC AAH33714;  
XX  
DT 03-SEP-2001 (first entry)  
XX  
DE Human colon cancer antigen encoding cDNA SEQ ID NO:770.  
XX  
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;  
KW colorectal carcinoma; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200122920-A2.  
XX  
PD 05-APR-2001.  
XX  
PF 28-SEP-2000; 2000WO-US26524.  
XX  
PR 29-SEP-1999; 99US-0157137.  
PR 03-NOV-1999; 99US-0163280.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ruben SM, Barash SC, Birse CE, Rosen CA;  
XX WPI; 2001-235357/24.  
DR P-PSDB; AAG74283.  
XX  
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
XX useful for preventing, diagnosing and/or treating colorectal cancers -  
PS Claim 1; Page 2745-2746; 9803pp; English.

XX  
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
CC the proteins are collectively known as colon cancer antigens. The colon  
CC cancer antigens have cytostatic activity and can be used in gene  
CC therapy and vaccine production. N and P may be used in the prevention,  
CC diagnosis and treatment of diseases associated with inappropriate P  
CC expression. For example, N and P may be used to treat disorders  
CC associated with decreased expression by rectifying mutations or deletions  
CC in a patient's genome that affect the activity of P by expressing  
CC inactive proteins or to supplement the patients own production of P.  
CC Additionally, N may be used to produce the colon cancer-associated Ps,  
CC by inserting the nucleic acids into a host cell and culturing the cell  
CC to express the proteins. N and P can be used in the prevention, diagnosis  
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
CC and AAH77789 represent sequences used in the exemplification of the  
CC present invention.  
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
CC missing at time of publication, meaning no sequences are present for  
CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
XX  
SQ Sequence 2468 BP; 753 A; 508 C; 466 G; 733 T; 8 other;  
Query Match 64.6%; Score 2316.8; DB 22; Length 2468;  
Best Local Similarity 95.4%; Pred. No. 0;  
Matches 2432; Conservative 8; Mismatches 0; Indels 110; Gaps 1;  
QY 1036 aggaaggacatcctcttctacacccagatggttagccacagggggcgttttgc 1095  
|||||  
Db 2 aaggaaggacatcctcttctacacccagatggttagccacagggggcgttttgc 61  
QY 1096 ctcttgcctgtgagaatatcacagaatgcttgagaagcataaagacttgttgggt 1155  
|||||  
Db 62 ctcttgcctgtgagaatatcacagaatgcttgagaagcataaagacttgttgggt 121  
QY 1156 aaagatacattgaactcttcaggagcacagcagctgaactcagcaggttgcgaatcga 1215  
|||||  
Db 122 aaagatacattgaactcttcaggagcacagcagctgaactcagcaggttgcgaatcga 181  
QY 1216 ttctctcgccctctcattccacttccaaacccctccacttattccactactacacag 1275  
|||||  
Db 182 ttctctcgccctctcattccacttccaaacccctccacttattccactactacacag 241  
QY 1276 caatttgcctccctacaaatgttagagctgtatcgccttcgaggtcttccctatgca 1335  
|||||  
Db 242 caatttgcctccctacaaatgttagagctgtatcgccttcgaggtcttccctatgca 301  
QY 1336 gccacaattgagacacatcctggatttctctggggaggttcgcccacagatatctactcat 1395  
|||||  
Db 302 gccacaattgagacacatcctggatttctctggggaggttcgcccacagatatctactcat 361  
QY 1396 ggggttcacatggttttgatcacccagggccgcccacatcaggagatgcctttatccagatg 1455  
|||||  
Db 362 ggggttcacatggttttgatcacccagggccgcccacatcaggagatgcctttatccagatg 421  
QY 1456 aagctcgcggacagacattttatgctgcacagaagtgtcataaaaaaacatgaaggac 1515  
|||||  
Db 422 aagctcgcggacagacattttatgctgcacagaagtgtcataaaaaaacatgaaggac 481  
QY 1516 agatatgttgaagtcttccagtggttcagctgagagatgaactttgtttaatggggggc 1575  
|||||  
Db 482 agatatgttgaagtcttccagtggttcagctgagagatgaactttgtttaatggggggc 541  
QY 1576 actttaaatcgaaatggtttatccccacgcgatgctgtctctccctccctcacattt 1635  
|||||  
Db 542 actttaaatcgaaatggtttatccccacgcgatgctgtctctccctccctcacattt 601  
QY 1636 ccagctcctgcgtcaattattctacagaagctgcatttacacagcctctgtgatttg 1695  
|||||  
Db 602 ccagctcctgcgtcaattattctacagaagctgcatttacacagcctctgtgatttr 661  
QY 1696 aatccacagcactgcagccctccacagcgtactaccacagcaggcactcagcttctcatg 1755



Db 662 aatccagagcactgcagccctccacagctactaccagcaggcactcagctctcatg 721  
Qy 1756 aactacacagcgtactatccacagccccaggttcgcctaaatagctctggctacttccct 1815  
Db 722 aactacacagcgtactatccacagccccaggttcgcctaaatagctctggctacttccct 781  
Qy 1816 acagctgctaactcttagcgggtgcctccacagcctgcacggtggcagaaatgcagggc 1875  
Db 782 acagctgctaactcttagcgggtgcctccacagcctgcacggtggcagaaatgcagggc 841  
Qy 1876 ctggcctcaataactggaggttaagaaattcttaactctctccaaggttaccagtatgca 1935  
Db 842 ctggcctcaataactggaggttaagaaattcttaactctctccaaggttacc----- 893  
Qy 1936 accgaggtgacttatcacacaaaatgaccagggcaggaactctaccacaaagaatgggtt 1995  
Db 894 ----- 893  
Qy 1996 tgtatttaaggccccagcagttagaacatcctcagaaaaagaagtgtttgaaagatgtat 2055  
Db 894 -----agtgtttgaaagatgtat 911  
Qy 2056 ggtgatcttgaaacctccagacacaaagaaactcttagcaaatccagggggaagtttgtct 2115  
Db 912 ggtgatcttgaaacctccagacacaaagaaactcttagcaaatccagggggaagtttgtct 971  
Qy 2116 acactcagcgtgcagtatcttcagcaaaacttgattggcacaacggcgtgaccttatct 2175  
Db 972 acactcagcgtgcagtatcttcagcaaaacttgattggcacaacggcgtgaccttatct 1031  
Qy 2176 ttggtgagtgaaaaatttgagctagtgaaagccaaatcgtaacttacacaaagcagca 2235  
Db 1032 ttggtgagtgaaaaatttgagcyagtgaaagccaaatcgtaacttacacaaagcagca 1091  
Qy 2236 tgcagcatacctggtcttctgattgcataataggaacttaaaatgtgaatttggaaatc 2295  
Db 1092 tgcagcatacctggtcttctgattgcataataggaacttaaaatgtgaatttggaaatc 1151  
Qy 2296 agatgtcccatctacttcacgttaaagtggcatcataggtgttccctaaagttttaaagctc 2355  
Db 1152 agatgtcccatctacttcacgttaaagtggcatcataggtgttccctaaagttttaaagctc 1211  
Qy 2356 tggataaaaactccaccagtgctaccatctccacatcgaaactcgtttaaagaaagcttca 2415  
Db 1212 tggataaaaactccaccagtgctaccatctccacatcgaaactcgtttaaagaaagcttca 1271  
Qy 2416 ttttggatatcccgctcttctctcatttccctgctcttctgcataatcagctcttc 2475  
Db 1272 ttttggatatcccgctcttctctcatttccctgctcttctgcataatcagctcttc 1331  
Qy 2476 ttgctaagtaattcaagcataagactcttggataataataaaatcacaaactcttaggagaaga 2535  
Db 1332 ttgctaagtaattcaagcataagactcttggataataataaaatcacaaactcttaggagaaga 1391  
Qy 2536 ataaaattgttatcttccagctctcttggccatgatgatataatttattgattaaaaacaaat 2595  
Db 1392 ataaaattgttatcttccagctctcttggccatgatgatataatttattgattaaaaacaaat 1451  
Qy 2596 faaaattttaaacaactgaagataaataatagaaaattgtgcacccctccacaaaacatac 2655  
Db 1452 faaaattttaaacaactgaagataaataatagaaaattgtgcacccctccacaaaacatac 1511  
Qy 2656 aaagtgttaaaagtgttgatcttctcagcaggttatcagttgttaataatgaattaggg 2715  
Db 1512 aaagtgttaaaagtgttgatcttctcagcaggttatcagttgttaataatgaattaggg 1571  
Qy 2716 gccaaaaatgcacaaacgaaaaatgaagcagctacatgtgttagttgaatttcttagttgaac 2775  
Db 1572 gccaaaaatgcacaaacgaaaaatgaagcagctacatgtgttagttgaatttcttagttgaac 1631  
Qy 2776 tgaattgaattgtggtctcatatgtattatttatattgtactttttcattattga 2835

Db 1632 tgytaattgaatatttggttcctteatgtattattttatatttatttatttatttatttga 1691  
Qy 2836 tggtttgactttaaagagaaaattcccatagtttttaataatccagaaagtgcagacaatt 2895  
Db 1692 tggtttgactttaaagagaaaattcccatagtttttaataatccagaaagtgcagacaatt 1751  
Qy 2896 tgaacagtgattcttagaaaaaatcacactaaactgaacagaagtgaatgcttatatatat 2955  
Db 1752 tgaacagtgattcttagaaaaaatcacactaaactgaacagaagtgaatgcttatatatat 1811  
Qy 2956 tatgatacctttaaacctttttctctaatgcttaactgaactgaataataataataaactttt 3015  
Db 1812 tatgatacctttaaacctttttctctaatgcttaactgaactgaataataataataaactttt 1871  
Qy 3016 aaagcataggactatagtcagcatgctagactgagaggttaaacactgatgcaattagaac 3075  
Db 1872 aaagcataggactatagtcagcatgctagactgagaggttaaacactgatgcaattagaac 1931  
Qy 3076 aggtactgatgctgcagtggtttaacactatgttttagctgtgtttatgctataaaagtgc 3135  
Db 1932 aggtactgatgctgcagtggtttaacactatgttttagctgtgtttatgctataaaagtgc 1991  
Qy 3136 aatattaacactagctagctagctgcctcatgtaactccaacaaacacagatttcat 3195  
Db 1992 aatattaacactagctagctagctgcctcatgtaactccaacaaacacagatttcat 2051  
Qy 3196 taagtgcattgaatgggatatttcttaagttaactcatattgtctcttgccttgaatgca 3255  
Db 2052 taagtgcattgaatggmtatttcttaagttaactcatattgtctcttgccttgaatgca 2111  
Qy 3256 atgcogtgcagatttatgagcgtgctattttattttctgtgctacttaacacctta 3315  
Db 2112 atgcogtgcagatttatgwggtgctattttattttctgtcattacttaacacctta 2171  
Qy 3316 aaggagaagcaaacatttcttcagctgactggaatggccctttaactgcaatag 3375  
Db 2172 aaggagaagcaaacatttcttcagctgactggaatggccctttaactgcaatag 2231  
Qy 3376 gaagaaaaaaaggggttggtaaaattggtaactggcacttaagatcgaaaaag 3435  
Db 2232 gaagaaaaaaaggggttggtaaaattggtaactggcacttaagatcgaaaaag 2291  
Qy 3436 aaattctgtatacttgactgacttaagatgccccaaagctgccccaaagcttgaagactt 3495  
Db 2292 aaattctgtatacttgactgacttaagatgccccaaagctgccccaaagcttgaagactt 2351  
Qy 3496 taagatagcagtaatgcttactacaactactagttttttgttagagttaacatttgtat 3555  
Db 2352 taagatagcagtaatgcttactacaactactagttttttgttagagttaacatttgtat 2411  
Qy 3556 aataaaacttgctgttttaactcaaaaa 3585  
Db 2412 aataaaacttgctgttttaactcaaaaa 2441  
RESULT 5  
ID AAC77930 standard; cDNA; 2261 BP.  
XX AAC77930;  
AC AAC77930;  
XX AAC77930;  
DT 08-FEB-2001 (first entry)  
XX  
DE Human cancer associated gene sequence SEQ ID NO:324.  
XX  
KW Human; cancer associated gene; cancer antigen; detection; cancer;  
KW diagnosis; cytostatic; proliferative; antineoplastic; immunomodulator;  
KW antidiabetic; antischistosomal; antineoplastic; antithrombotic; antiviral;  
KW antiinflammatory; antihypertensive; antihyperlipidemic; antitumor;  
KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;  
KW vasotrophic; antipsoriatic; antiangiogenic; gene therapy; inflammation;  
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;  
KW allergic reaction; graft versus host disease; organ rejection;





QY 2442 ttcatctccctgtcttctgcataatcatgctcttcttctgtaagtaattcaagcataagatc 2501  
Db 1120 ttcatctccctgtcttctgcataatcatgctcttcttctgtaagtaattcaagcataagatc 1179  
QY 2502 ttggaataataaaatcaacaatcttaggagaagaataaaattgtattttccagctctct 2561  
Db 1180 ttggaataataaaatcaacaatcttaggagaagaataaaattgtattttccagctctct 1239  
QY 2562 tggccatgatgatctcttatgattaaacacaaataataatttttaaaacacctgaagataaa 2621  
Db 1240 tggccatgatgatctcttatgattaaacacaaataataatttttaaaacacctgaagataaa 1299  
QY 2622 ttagaagaattgtgcacccctccacaaacacatacaaaagttttaaaagtgtgatcttttc 2681  
Db 1300 ttagaagaattgtgcacccctccacaaacacatacaaaagttttaaaagtgtgatcttttc 1359  
QY 2682 tcagcaggtatcagttgtaataatgaattagggccaaaatgcaaaacgaaaatgaag 2741  
Db 1360 tcagcaggtatcagttgtaataatgaattagggccaaaatgcaaaacgaaaatgaag 1419  
QY 2742 cagctcacatgtagtttagtaattcttagttgaactgttaattgaattatgtggtctcatat 2801  
Db 1420 cagctcacatgtagtttagtaattcttagttgaactgttaattgaattatgtggtctcatat 1479  
QY 2802 gtattatttatattgtactttttcttcatattatgtatggtttggacctttaataagaaaatt 2861  
Db 1480 gtattatttatattgtactttttcttcatattatgtatggtttggacctttaataagaaaatt 1539  
QY 2862 coatagttttaataatccagagtgagacaatttgaacagtgtaattcttagaaaacaata 2921  
Db 1540 coatagttttaataatccagagtgagacaatttgaacagtgtaattcttagaaaacaata 1599  
QY 2922 cactaacgtgaacagagtgaaatgcttatataattatgatagccttaacctttttctctc 2981  
Db 1600 cactaacgtgaacagagtgaaatgcttatataattatgatagccttaacctttttctctc 1659  
QY 2982 taatgccttaactgtcaataatataacttttaaacctataggactatagtcagcatgac 3041  
Db 1660 taatgccttaactgtcaataatataacttttaaacctataggactatagtcagcatgac 1719  
QY 3042 tagactgagagtgaaacactgtatgaatttagaacaggtactgtatgctgcagtggttaac 3101  
Db 1720 tagactgagagtgaaacactgtatgaatttagaacaggtactgtatgctgcagtggttaac 1779  
QY 3102 actatgttttagctgtgtttatgctataaaagtgcataattatagacactagctagtactgt 3161  
Db 1780 actatgttttagctgtgtttatgctataaaagtgcataattatagacactagctagtactgt 1839  
QY 3162 gctcatgttaactcaagagaacacaggatttcatatgaattgaattgaattgataattttc 3221  
Db 1840 gctcatgttaactcaagagaacacaggatttcatatgaattgaattgaattgataattttc 1899  
QY 3222 ctaagtactcatatgtctcttctgtgaatgcaatgcctgacagatttatgagctgtct 3281  
Db 1900 ctaagtactcatatgtctcttctgtgaatgcaatgcctgacagatttatgagctgtct 1959  
QY 3282 attttttttttctgtcattactttaaacacctttaaaggagagagaacacattctctctt 3341  
Db 1960 attttttttttctgtcattactttaaacacctttaaaggagagagaacacattctctctt 2019  
QY 3342 cagctgactggaatggccctttaactgcaataggaagaagaaaaaagggtttgtgtga 3401  
Db 2020 cagctgactggaatggccctttaactgcaataggaagaagaaaaaagggtttgtgtga 2079  
QY 3402 aaatgggtgataactggcacttaagatgcagaagaataattctgtactatgctgacttaa 3461  
Db 2080 aaatgggtgataactggcacttaagatgcagaagaataattctgtactatgctgacttaa 2139  
QY 3462 gatgcccagaagctgccaaagctctgaaagactttaagatagcgagtaatgcttactaca 3521  
Db 2140 gatgcccagaagctgccaaagctctgaaagactttaagatagcgagtaatgcttactaca 2199  
QY 3522 atactactgagttttttgtgagagttaacatttgcataataaaactgctgtttaactcaaa 3581

Db 2200 atactactgagttttgtagagtttaacatttgataataaaactgctgtttaactcaaa 2259  
QY 3582 aa 3583  
Db 2260 aa 2261  
RESULT 6  
AAH14740  
ID AAH14740 standard; cDNA: 1624 BP.  
XX AAH14740;  
AC AAH14740;  
XX 26-JUN-2001 (first entry)  
XX Human cDNA sequence SEQ ID NO:12485.  
XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.  
XX Homo sapiens.  
XX EP1074617-A2.  
XX 07-FEB-2001.  
PD 28-JUL-2000; 2000EP-0116126.  
PF 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI: 2001-318749/34.  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
full-length cDNAs defined in the specification, and for the detection  
of the full-length cDNAs -  
XX full-length cDNAs -  
PS Claim 8; SEQ ID 12485; 2537pp + CD ROM; English.  
XX The present invention describes primer sets for synthesizing 5602  
full-length cDNAs defined in the specification. Where a primer set  
comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
to the complementary strand of a polynucleotide which comprises one of  
the 5602 nucleotide sequences defined in the specification, where the  
oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
of an oligonucleotide comprising a sequence complementary to the  
complementary strand of a polynucleotide which comprises a 5'-end  
sequence and an oligonucleotide comprising a sequence complementary to a  
polynucleotide which comprises a 3'-end sequence, where the  
oligonucleotide comprises at least 15 nucleotides and the combination of  
the 5'-end sequence/3'-end sequence is selected from those defined in  
the specification. The primer sets can be used in antisense therapy and  
in gene therapy. The primers are useful for synthesizing polynucleotides,  
particularly full-length cDNAs. The primers are also useful for the  
detection and/or diagnosis of the abnormality of the proteins encoded by  
the full-length cDNAs. The primers allow obtaining of the full-length  
cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
AAH13633 to AAH18742 represent human cDNA sequences. AAB92446 to  
AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
represent oligonucleotides, all of which are used in the exemplification  
of the present invention.  
XX Sequence 1624 BP; 498 A; 320 C; 288 G; 518 T; 0 other;

Query Match	37.0%	Score 1324.8	DB 22	Length 1624	
Best Local Similarity	99.5%	Pred. No. 0			
Matches 1329	Conservative 0	Mismatches 7	Indels 0	Gaps 0	
QY	2038	agtggttgaagaagatgtatgtgatcttgaacatccagacacaagaataactcttagcaaa	2097		
Db	289	agtgcttgaagaagtatgtgacttgaacacctccagacacaagaataactcttagcaaa	348		
QY	2098	tccaggggaagttgtctacactcaggctgcagtattttcagcaaaccttgattggacaaa	2157		
Db	349	tccaggggaagttgtctacactcaggctgcagtattttcagcaaaccttgattggacaaa	408		
QY	2158	cgggcctgtgccttatcttcttgggtggagtgaaaaaatttgagctagtgaaagccaaatcgt	2217		
Db	409	cgggcctgtgccttatcttcttgggtggagtgaaaaaatttgagccagtgaaagccaaatcgt	468		
QY	2218	aacttacagaacgacgacatcagcatcaccctgcgtctcttgcgtgattgaaataggcattta	2277		
Db	469	aacttacagaacgacgacatcagcatcaccctgcgtctcttgcgtgattgaaataggcattta	528		
QY	2278	aaatgtgaatttggaaatcagatgtctccatttaacttcacgttcaagtgaggcatcagggtgt	2337		
Db	529	aaatgtgaatttggaaatcagatgtctccatttaacttcacgttcaagtgaggcatcagggtgt	588		
QY	2338	ttctcaagttttaaagttcttgataaaaaactccaccagtgctaccatctccaccatgaac	2397		
Db	589	ttctcaagttttaaagttcttgataaaaaactccaccagtgctaccatctccaccatgaac	648		
QY	2398	tcgtttaaggaagcttcatttttgtatattccgcgtctttctctcttaattccctgtcttt	2457		
Db	649	tcgtttaaggaagcttcatttttgtatattccgcgtctttctctcttaattccctgtcttt	708		
QY	2458	ctgcataatcatgctctctgtctaagttaattccaagcataagatcttggaaataataaaatc	2517		
Db	709	ctgcataatcatgctctctgtctaagttaattccaagcataagatcttggaaataataaaatc	768		
QY	2518	acaactcttaggagaaagataaaaattgtatttccccagtcctcttggccatgatgatc	2577		
Db	769	acaactcttaggagaaagataaaaattgtatttccccagtcctcttggccatgatgatc	828		
QY	2578	ttatgattaaaaacaaatttaaaattttaaaccacctgaagataaaattagaagaatttgc	2637		
Db	829	ttatgattaaaaacaaatttaaaattttaaaccacctgaagataaaattagaagaatttgc	888		
QY	2638	accctccacaaaacatacacaagttttaaagtttggatctttttctccagcaggtatcagtt	2697		
Db	889	accctccacaaaacatacacaagttttaaagtttggatctttttctccagcaggtatcagtt	948		
QY	2698	gtaataatgaattaggggccaaaatgcataaagtcgaaagtcgacgtacatgattta	2757		
Db	949	gtaataatgaattaggggccaaaatgcataaagtcgaaagtcgacgtacatgattta	1008		
QY	2758	gtaatttctagtttgaactgttaattgtaatttgygcttccatattgtatttttatattg	2817		
Db	1009	gtaatttctagtttgaactgttaattgtaatttgygcttccatattgtatttttatattg	1068		
QY	2818	tactttttcatttatgattggttggacttttaataagagaataattcccatagttttaatat	2877		
Db	1069	tactttttcatttatgattggttggacttttaataagagaataattcccatagttttaatat	1128		
QY	2878	cccagaagtgagacaaattgaacagtgtaattctagaaaaacaatacactaaactgaacagaa	2937		
Db	1129	cccagaagtgagacaaattgaacagtgtaattctagaaaaacaatacactaaactgaacagaa	1188		
QY	2938	gtgaattccttatatatattatgatagccttaaacctttttctccttaatgccttaactgtc	2997		
Db	1189	gtgaattccttatatatattatgatagccttaaacctttttctccttaatgccttaactgtc	1248		
QY	2998	aaataaattaaaccttttaagacataggaactatagtcagcatgctagactgagaggtaaa	3057		
Db	1249	aaataaattaaaccttttaagacataggaactatagtcagcatgctagactgagaggtaaa	1308		

CC nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d)  
CC wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
XX Sequence 1585 BP; 385 A; 353 C; 345 G; 502 T; 0 other;

Query Match 16.4%; Score 587.4; DB 21; Length 1585;  
Best Local Similarity 99.3%; Pred. No. 5.9e-141;  
Matches 611; Conservative 0; Mismatches 1; Indels 3; Gaps 2;

QY 1 cttttggatcactgctggggccaccgggcccagctaggctcgatgagaaggagt 60  
DB 155 cttttggatcactgctggggccaccgggcccagctaggctcgatgagaaggagt 214  
QY 61 atctctgttctggaagtctggtatctggccacaagaagtgaggacagttgcacga 120  
DB 215 atctctgttctggaagtctggtatctggccacaagaagtgaggacagttgcacga 274  
QY 121 gtctagttagaccgcatcagttggaactgacgagagactgcaaaagaagaactaaata 180  
DB 275 --tctagttagaccgcatcagttggaactgacgagagactgcaaaagaagaactaaata 332  
QY 181 gacgtcgaaagcctctcctcgctgcagctggaccagccctccgacagtttaaccag 240  
DB 333 gacgtcgaaagcctctcctcgctgcagctggaccagccctccgacagtttaaccag 392  
QY 241 tcagtgagcaatgaactgaattgagtaggagacttctctgtctgtactgat-gg 299  
DB 393 tcagtgagcaatgaactgaattgagtaggagacttctctgtctgtactgatgg 452  
QY 300 gcaagctcatgacagcaaatcctgcatcctgaggtctccaaagaagaatgtactattacc 359  
DB 453 gcaagctcatgacagcaaatcctgcatcctgaggtctccaaagaagaatgtactattacc 512  
QY 360 tgaatgcttctattctctttttgatcttcgaaagaattcaagaagaatgtgacctggttc 419  
DB 513 tgaatgcttctattctctttttgatcttcgaaagaattcaagaagaatgtgacctggttc 572  
QY 420 acctgatattgacaaactggaacttgcacaaatgacagagatttttaaatgtgagaagag 479  
DB 573 acctgatattgacaaactggaacttgcacaaatgacagagatttttaaatgtgagaagag 632  
QY 480 tagttcagttctcgatagagcctctcaagttagagatatgggaataataattttagc 539  
DB 633 tagttcagttctcgatagagcctctcaagttagagatatgggaataataattttagc 692  
QY 540 aatgatttcagagccttataataacacaggttttcagatccagagagagtggaattacaagt 599  
DB 693 aatgatttcagagccttataataacacaggttttcagatccagagagagtggaattacaagt 752  
QY 600 tgaagtggaaacttg 614  
DB 753 tgaagtggaaacttg 767

RESULT 8  
AAZ16154  
ID AAZ16154 standard; cDNA; 740 BP.

XX AAZ16154;

XX AC  
XX 12-OCT-1999 (first entry)

XX DE Human gene expression product cDNA sequence SEQ ID NO:3624.

XX Human; gene: gene expression product; diagnosis; therapy; probe;  
KW detection; mapping; tissue typing; profiling; forensic; cancer;  
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.  
XX Homo sapiens.  
PN WO9938972-A2.  
PD 05-AUG-1999.  
XX 28-JAN-1999; 99WO-US01619.  
XX 03-APR-1998; 98US-0080666.  
PR 28-JAN-1998; 98US-0072910.  
PR 24-FEB-1998; 98US-0075954.  
PR 31-MAR-1998; 98US-0080114.  
PR 03-APR-1998; 98US-0080515.  
XX (CHIR ) CHIRON CORP.  
PA (HYSE-) HYSEQ INC.  
XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;  
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;  
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;  
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;  
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;  
XX WPI; 1999-494092/41.  
XX Novel human genes and their expression products which are  
PT differentially expressed in different cell types  
XX Claim 1; Page 1729; 2479pp; English.  
XX The present invention describes a library of human polynucleotides  
CC comprising the sequences given in AAZ12532 to AAZ17799. Also described is  
CC a method of detecting differentially expressed genes correlated with the  
CC cancerous state of a mammalian cell, comprising detecting at least one  
CC differentially expressed gene product in a test sample from a cell  
CC suspected of being cancerous, where the gene product is encoded by one  
CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17799. The  
CC polynucleotides can be used as a source of primers and probes, which can  
CC be used for a variety of purpose, e.g. detection of expression levels,  
CC mapping, tissue typing or profiling, forensics, genetic analysis and  
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides  
CC can be used for raising antibodies for experimental, diagnostic and  
CC therapeutic purposes. The polynucleotides may also be used to construct  
CC arrays for diagnostics (which may be used to determine function of an  
CC encoded protein); and to detect differences in expression levels between  
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to  
CC identify a genetic predisposition or susceptibility to a disease such as  
CC cancer). The polynucleotides of the invention are especially used in the  
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,  
CC and lung cancer. The polynucleotides can also be used to screen for  
CC peptide analogues and antagonists.  
XX Sequence 740 BP; 188 A; 163 C; 183 G; 193 T; 13 other;

Query Match 15.5%; Score 555.6; DB 20; Length 740;  
Best Local Similarity 98.5%; Pred. No. 6.4e-133;  
Matches 591; Conservative 0; Mismatches 6; Indels 3; Gaps 3;

QY 1 cttttggatcactgctggggccaccgggcccagctaggctcgatgagaaggagt 60  
DB 143 cttttggatcactgctggggccaccgggcccagctaggctcgatgagaaggagt 202  
QY 61 atctctgttcttggaaagtcgtggatctggccacaagaagtgaggacagttgcacga 120  
DB 203 atctctgttcttggaaagtcgtggatctggccacaagaagtgaggacagttgcacga 262  
QY 121 gtgctagttagaccgcatcagttggaactgacgagagactgcaaaagaagaactaaata 180



Db 744 tt 745

RESULT 10

AAH12166/C

ID AAH12166 standard; cDNA; 565 BP.

XX

AC AAH12166;

XX

DT 26-JUN-2001 (first entry)

XX

DE Human cDNA clone (3'-primer) SEQ ID NO:9001.

XX

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX

OS Homo sapiens.

XX

PN EP1074617-A2.

XX

PD 07-FEB-2001.

XX

PF 28-JUL-2000; 2000EP-0116126.

XX

PR 29-JUL-1999; 99JP-0248036.

PR

PR 27-AUG-1999; 99JP-0300253.

PR

PR 11-JAN-2000; 2000JP-0118776.

PR

PR 02-MAY-2000; 2000JP-0183767.

PR

PR 09-JUN-2000; 2000JP-0241899.

XX

PA (HELI-) HELIX RES INST.

XX

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX

DR WPI: 2001-318749/34.

XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

XX

PS Claim 3; SEQ ID 9001; 2537pp + CD ROM; English.

XX

CC The present invention describes primer sets for synthesizing 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

XX

XX Sequence 565 BP; 192 A; 97 C; 102 G; 172 T; 2 other;

XX

Query Match 14.1%; Score 505.4; DB 22; Length 565;

Best Local Similarity 95.8%; Pred. No. 4.8e-120;

Matches 541; Conservative 0; Mismatches 21; Indels 3; Gaps 2;

Qy 2810 ttatttgtaacttttttcaattgattgaggttggaactttaataagagaaattccaagt 2869

Db 563 TTTAAATGGACTTTTTCATAAT--GATGGTTGGACTTTTAATAGAGAATTCCCATAGTT 506

Qy 2870 ttaatatccagagaagtgagacaatttgacacagtgattctagagaaacaatacactaac- 2928

Db 505 TTTAATATCCAGAGTGAGCCATTTGAACAGGTGATTTCTAGAAACCAATACACTAACT 446

Qy 2929 tgaacagaagtgaatgcttatatatattatgatagccttaaaccttttcccttaagtcc 2988

Db 445 TGACCAGAGTGAATGCTTATATATATATATATATATATATATATATATATATAT 386

Qy 2989 ttaactgtcaaatataataacacctttaaacgactagggactatagtcagcagctagactg 3048

Db 385 TTAACCTGTCAAAATAATTATAAACCTTTTAAAGCATAGGACTATAGTCAGCATGCTAGACTG 326

Qy 3049 agaggtaaacactgatgcaatttagaacaaggtaactgactgactgactgactgactgact 3108

Db 325 AGAGGTAACACTGATGCAATTTAGAACAGGTACTGATGCTGCTGCTGCTGCTGCTGCT 266

Qy 3109 ttactgtgtttatgctataaaagtgcgaattatagacactagctagctagctagctagct 3168

Db 265 TTACTGTGTTTATGCTTATAAAGTGCATATATATATATATATATATATATATATAT 206

Qy 3169 gtaactccaaagaaacagagatttcatttaagtgcattgaatgtggatattctctaaagt 3228

Db 205 GTAACCTCCAAAGAAACAGAGATTTCATTAAGTGCATTTGAATGTGGCTATTTCTCTAAGTT 146

Qy 3229 actcatattgctcttctgctgtaagtgcgaatgccgtgcagatttatgagcgtctatttta 3288

Db 145 ACTCATATTTGCTTTTTCCTTGAATGCAATGCCGTGCAATTTATGCTGCTGCTATTTT 86

Qy 3289 ttctctgtgcttacttttaacacaccttaagggagaagcaaacatttctctcagctga 3348

Db 85 TTTTCTGTGCATTAATTAACACCTTAAAGGGAGAGCAACATTCCTTTCTTCAGCTGA 26

Qy 3349 ctggcaatggccctttaactgcaat 3373

Db 25 CTGGCAATGGCCCTTTAACTGCAAT 1

RESULT 11

AAF68133

ID AAF68133 standard; cDNA; 503 BP.

XX

AC AAF68133;

XX

DT 12-APR-2001 (first entry)

XX

DE Human lung tumour protein related nucleotide sequence SEQ ID NO:51.

XX

KW Human; lung cancer; lung tumour; lung tumour protein; gene therapy;

KW lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;

XX

OS Homo sapiens.

XX

PN WO200100828-A2.

XX

PD 04-JAN-2001.

XX

PF 30-JUN-2000; 2000WO-US18061.

XX

PR 30-JUN-1999; 99US-0346492.

PR

PR 15-OCT-1999; 99US-0419356.

PR

PR 17-DEC-1999; 99US-0466867.

PR

PR 30-DEC-1999; 99US-0476300.

PR

PR 06-MAR-2000; 2000US-0519642.

PR

PR 22-MAR-2000; 2000US-0533077.

PR

PR 10-APR-2000; 2000US-0546259.

PR

PR 27-APR-2000; 2000US-0560406.

PR

PR 05-JUN-2000; 2000US-0589184.

XX PA (CORI-) CORIXA CORP.  
XX PI Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;  
XX PI Retter MW, Mannion J;  
XX XX  
XX DR WPI; 2001-071488/08.  
XX XX  
XX PT Lung tumor-associated proteins and the nucleic acids that encode them,  
XX PT useful for preventing, diagnosing and treating lung cancer -  
XX XX  
XX PS Claim 4; Page 168; 436pp; English.  
XX XX  
XX CC The present invention describes immunogenic portions of lung tumour-  
XX CC associated proteins (I) and the nucleic acids (NAs) that encode them.  
XX CC (I) have cytostatic activity and can be used in gene therapy, antisense  
XX CC inhibition and in vaccines. The NAs and the lung tumour-associated  
XX CC proteins they encode may be used in the prevention, treatment and  
XX CC diagnosis of diseases associated with their inappropriate expression,  
XX CC especially lung cancers. For example, the NAs may be administered to  
XX CC treat diseases by rectifying mutations or deletions in a patient's genome  
XX CC that affect the activity of the protein by expressing inactive proteins  
XX CC or to supplement the patients own production of (I). Additionally, the  
XX CC NAs may be used to produce the lung-tumour associated protein, according  
XX CC to standard recombinant DNA methodology. Conversely, antisense NA  
XX CC molecules may be administered to down regulate protein expression by  
XX CC binding with the cells own genes and preventing their expression. The NA  
XX CC and complementary sequences may also be used as DNA probes in diagnostic  
XX CC assays to detect and quantitate the presence of similar NA sequences in  
XX CC samples, and hence which patients may be in need of treatment for lung  
XX CC cancer. The (I) may be used as antigens in the production of antibodies  
XX CC and in assays to identify modulators (agonists and antagonists) of the  
XX CC expression and activity of the protein. AAF68083 to AAF68878 and  
XX CC AAB76848 to AAB76878 represent human lung tumour protein related  
XX CC nucleotide and protein sequences which are used in the exemplification  
XX CC of the present invention.  
XX XX  
XX S0 Sequence 503 BP; 188 A; 69 C; 78 G; 168 T; 0 other;

Query Match 13.98; Score 498.4; DB 22; Length 503;  
Best Local Similarity 99.88; Pred. No. 2.9e-118;  
Matches 499; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2572 gatacttattgataaaacaaataattttaaaacacctgaagataaaattagaagaaa 2631  
Db 1 gatacttattgataaaacaaataattttaaaacacctgaagataaaattagaagaaa 60  
Qy 2632 ttgtgacccctcccaaaacatacaaaagttaaagtttaaagtttgatcttttctcagcaggtg 2691  
Db 61 ttgtgacccctcccaaaacatacaaaagttaaagtttgatcttttctcagcaggtg 120  
Qy 2692 tcagttgataaataaataatgagggccaaataatgcaaaacgaaataaagaagcagctacatg 2751  
Db 121 tcagttgataaataaataatgagggccaaataatgcaaaacgaaataaagaagcagctacatg 180  
Qy 2752 tagttagtaattcttagttgaactgaataattggaattgtgcttcattatatttt 2811  
Db 181 tagttagtaattcttagttgaactgaataattggaattgtgcttcattatatttt 240  
Qy 2812 atattgtactttttcattattgattggtttggaactttaataagaagaattccatagtttt 2871  
Db 241 atattgtactttttcattattgattggtttggaactttaataagaagaattccatagtttt 300  
Qy 2872 taatatccagaagtgaacaaatttgaacagtgatttcttagaataaacaatacaactaactga 2931  
Db 301 taatatccagaagtgaacaaatttgaacagtgatttcttagaataaacaatacaactaactga 360  
Qy 2932 acagaagtgaatgcttatatatattatgatagccttaaacctttttctcctaagcctta 2991  
Db 361 acagaagtgaatgcttatatatattatgatagccttaaacctttttctcctaagcctta 420  
Qy 2992 actgtcaataattataaaccttttaaaagcataggactatagtcagcatgctagactgaga 3051

Db 421 actgtcaataattataaaccttttaaaagcataggactatagtcagcatgctagactgaga 480  
Qy 3052 ggttaaacactgatgaattta 3071  
Db 481 ggttaaacactgatgaattta 500  
RESULT 12  
AAF68134  
ID AAF68134 standard; cDNA; 503 BP.  
XX  
XX AAF68134;  
XX  
XX 12-APR-2001 (first entry)  
XX  
XX Human lung tumour protein related nucleotide sequence SEQ ID NO:52.  
XX  
XX Human; lung cancer; lung tumour; lung tumour protein; gene therapy;  
XX KW lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;  
XX KW cytostatic; antisense inhibition; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200100828-A2.  
XX  
XX 04-JAN-2001.  
XX  
XX 30-JUN-2000; 2000WO-US18061.  
XX  
XX 30-JUN-1999; 99US-0346492.  
XX PR 15-OCT-1999; 99US-0419356.  
XX PR 17-DEC-1999; 99US-0466867.  
XX PR 30-DEC-1999; 99US-0476300.  
XX PR 06-MAR-2000; 2000US-0519642.  
XX PR 22-MAR-2000; 2000US-0533077.  
XX PR 10-APR-2000; 2000US-0546259.  
XX PR 27-APR-2000; 2000US-0560406.  
XX PR 05-JUN-2000; 2000US-0589184.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;  
XX PI Retter MW, Mannion J;  
XX  
XX WPI; 2001-071488/08.  
XX  
XX Lung tumor-associated proteins and the nucleic acids that encode them,  
XX PT useful for preventing, diagnosing and treating lung cancer -  
XX XX  
XX PS Claim 4; Page 168; 436pp; English.  
XX  
XX The present invention describes immunogenic portions of lung tumour-  
XX CC associated proteins (I) and the nucleic acids (NAs) that encode them.  
XX CC (I) have cytostatic activity and can be used in gene therapy, antisense  
XX CC inhibition and in vaccines. The NAs and the lung tumour-associated  
XX CC proteins they encode may be used in the prevention, treatment and  
XX CC diagnosis of diseases associated with their inappropriate expression,  
XX CC especially lung cancers. For example, the NAs may be administered to  
XX CC treat diseases by rectifying mutations or deletions in a patient's genome  
XX CC that affect the activity of the protein by expressing inactive proteins  
XX CC or to supplement the patients own production of (I). Additionally, the  
XX CC NAs may be used to produce the lung-tumour associated protein, according  
XX CC to standard recombinant DNA methodology. Conversely, antisense NA  
XX CC molecules may be administered to down regulate protein expression by  
XX CC binding with the cells own genes and preventing their expression. The NA  
XX CC and complementary sequences may also be used as DNA probes in diagnostic  
XX CC assays to detect and quantitate the presence of similar NA sequences in  
XX CC samples, and hence which patients may be in need of treatment for lung  
XX CC cancer. The (I) may be used as antigens in the production of antibodies  
XX CC and in assays to identify modulators (agonists and antagonists) of the  
XX CC expression and activity of the protein. AAF68083 to AAF68878 and  
XX CC AAB76848 to AAB76878 represent human lung tumour protein related

QY 1577 ctttaaatcgaaatggcttatccaccacgccatgctgctcc 1619





